Proceedings of the Indian Academy of Sciences
Mathematical Sciences

Annual Subscription Rates

All countries except India (Price includes AIR MAIL charges)	US$ 100
India	Rs. 200

Annual subscriptions are available for **Individuals** for India and abroad at the concessional rates of Rs. 125/- and $ 30 respectively.

All correspondence regarding subscription should be addressed to **The Circulation Department** of the Academy.

Editorial Office

Indian Academy of Sciences, C V Raman Avenue,
P.B. No. 8005, Bangalore 560 080, India

Telephone: 80-334 2546, 80-334 2943
Telefax: 91-80-334 6094

Email: mathsci@ias.ernet.in

The "Notes on the preparation of papers" are printed in the last issue of every volume.

Proceedings of the Indian Academy of Sciences
Mathematical Sciences

Volume 110
2000

Published by the Indian Academy of Sciences
Bangalore 560 080

Proceedings of the Indian Academy of Sciences
Mathematical Sciences

Annual Subscription Rates

All countries except India (Price includes AIR MAIL charges)	US$ 100
India	Rs. 200

Annual subscriptions are available for **Individuals** for India and abroad at the concessional rates of Rs. 125/- and $ 30 respectively.

All correspondence regarding subscription should be addressed to **The Circulation Department** of the Academy.

Editorial Office

Indian Academy of Sciences, C V Raman Avenue,
P.B. No. 8005, Bangalore 560 080, India

Telephone: 80-334 2546, 80-334 2943
Telefax: 91-80-334 6094
Email: mathsci@ias.emet.in
Website: www.ias.ac.in/mathsci/

The "Notes on the preparation of papers" are printed in the last issue of every volume.

Proceedings of the Indian Academy of Sciences

Mathematical Sciences

Volume 110, 2000

VOLUME CONTENTS

Number 1, February 2000

Number 2, May 2000

Number 3, August 2000

Number 4, November 2000

Proc. Indian Acad. Sci. (Math. Sci.), Vol. 110, No. 1, February 2000, pp. 1–26.
© Printed in India

Intermediate Jacobians and Hodge structures of moduli spaces

DONU ARAPURA and PRAMATHANATH SASTRY*

Department of Mathematics, Purdue University, West Lafayette, IN 47907-1395, USA
*The Mehta Research Institute, Chhatnag, Jhusi, Allahabad 221 506, India
E-mail: dvb@math.purdue.edu; pramath@mri.ernet.in

MS received 21 April 1999; revised 4 August 1999

Abstract. The mixed Hodge structure on the low degree cohomology of the moduli space of vector bundles on a curve is studied. Analysis of the third cohomology yields a new proof of a Torelli theorem.

Keywords. Vector bundle; mixed Hodge structure; intermediate Jacobian.

1. Introduction

We work throughout over the complex numbers $\mathbb{C}$, i.e. all schemes are over $\mathbb{C}$ and all maps of schemes are maps of $\mathbb{C}$-schemes. A curve, unless otherwise stated, is a smooth complete curve. Points mean geometric points. We will, as is usual in such situations, toggle between the algebraic and analytic categories without warning.

For a curve X, $\mathcal{SU}_X(n,\, L)$ will denote the moduli space of *semi-stable* vector bundles of rank n and determinant L. The smooth open subvariety defining the *stable locus* will be denoted $\mathcal{SU}_X^s(n,\, L)$. We assume familiarity with the basic facts about such a moduli space as laid out, for example in [22] pp. 51–52, VI.A (see also Theorems 10, 17 and 18 of *loc. cit.*).

When L is a line bundle of degree coprime to n, the moduli spaces $\mathcal{SU}_X(n,L)$ and $\mathcal{SU}_X^s(n,L)$ coincide, and are therefore smooth and projective. The cohomology groups $H^i(\mathcal{SU}_X(n,L),\mathbb{Q})$ carry pure Hodge structures which can, in principle, be determined by using a natural set of generators [2] and relations [13] for the cohomology ring; we will say more about this later. When the degree of L is not coprime to n and $g > 2$, the situation is complicated by the fact that $\mathcal{SU}_X(n,L)$ is singular and $\mathcal{SU}_X^s(n,L)$ nonprojective. Thus the cohomology groups of these spaces carry (*a priori*) mixed Hodge structures, and it is these structures that we wish to understand. Our main results concerns the situation in low degrees.

Theorem 1.0.1. *Let* $\imath(n,g) = 2(n-1)g - (n-1)(n^2+3n+1) - 7$. *Let* X *be a curve of genus* $g \geq 2$. *If* $n \geq 4$ *and* $i < \imath(n,g)$ *are integers, then for any pair of line bundles* L, L' (*not necessarily with the same degree*) *on* X, *the mixed Hodge structures* $H^i(\mathcal{SU}_X^s(n,L),\mathbb{Q})$ *and* $H^i(\mathcal{SU}_X^s(n,L'),\mathbb{Q})$ *are* (*noncanonically*) *isomorphic and are both pure c weight* i.

This statement is a bit disingenuous, it is vacuous unless $g \geq 16$. The exp determination of these Hodge structures is rather delicate. However general consider tions show that these Hodge structures are semisimple and it is not difficult to write dc all the potential candidates for the simple summands.

COROLLARY 1.0.1

With the notation as above, for $i < \imath(n, g)$ any simple summand of

$$H^i(\mathcal{SU}^s_X(n, L), \mathbb{Q})$$

is, up to Tate twisting, a direct summand of a tensor power of $H^1(X)$.

For third cohomology, a more refined analysis yields:

Theorem 1.0.2. *Let X be a curve of genus $g \geq 2$, $n \geq 2$ an integer and L a line bundle on X. Let $\mathcal{S}^s = \mathcal{SU}^s_X(n, L)$.*

(a) *If $g > (3/(n-1)) + ((n^2+3n+3)/2)$ and $n \geq 4$, then $H^3(\mathcal{S}^s, \mathbb{Z})$ is a pure Hodge structure of type $\{(1,2), (2,1)\}$, and it carries a natural polarization making the intermediate Jacobian*

$$J^2(\mathcal{S}^s) = \frac{H^3(\mathcal{S}^s, \mathbb{C})}{F^2 + H^3(\mathcal{S}^s, \mathbb{Z})}$$

into a principally polarized abelian variety. There is an isomorphism of principally polarized abelian varieties $J(X) \simeq J^2(\mathcal{S}^s)$.

(b) *If $\deg L$ is a multiple of n, then the conclusions of (a) are true for $g \geq 3$, $n \geq 2$ except the case $g = 3$, $n = 2$.*

The word 'natural' above has the following meaning: an isomorphism between any two $\mathcal{S}^s$'s as above will induce an isomorphism on third cohomology which will respect the indicated polarizations. As an immediate corollary, we obtain the following Torelli theorem:

COROLLARY 1.0.2

Let X and X' be curves of genus $g \geq 3$, L and L' line bundles of (possibly different degrees) on X and X' respectively.

(a) *Assume that $n \geq 4$ is an integer such that $g > (3/(n-1)) + ((n^2+3n+1)/2)$. If*

$$\mathcal{SU}^s_X(n, L) \simeq \mathcal{SU}^s_{X'}(n, L') \tag{1.1}$$

or if

$$\mathcal{SU}_X(n, L) \simeq \mathcal{SU}_{X'}(n, L') \tag{1.2}$$

then

$$X \simeq X'.$$

(b) *If $\deg L = \deg L'$ and the common value is a multiple of n, then the conclusions of (a) are true for $n \geq 2$, except the case $g = 3$, $n = 2$.*

Proof. Since $\mathcal{SU}^s_X(n, L)$ (resp. $\mathcal{SU}^s_{X'}(n, L')$) is the smooth locus of $\mathcal{SU}_X(n, L)$ (resp. $\mathcal{SU}_{X'}(n, L')$), therefore it is enough to assume (1.1) holds. By assumption $J^2(\mathcal{SU}^s_X(n, L)) \simeq J^2(\mathcal{SU}^s_{X'}(n, L'))$ as polarized abelian varieties. Therefore $J(X) \simeq J(X')$, and the corollary follows from the usual Torelli theorem. □

When $(n, \deg L) = 1$ (the 'coprime case'), the second theorem (and its corollary with $\deg L = \deg L'$) has been proven by Narasimhan and Ramanan [18], Tyurin [24] (both in the range $n \geq 2$ and $g \geq 2$, except when $g = 2$, $n = 3$) and the special case of their results, when $n = 2$, by Mumford and Newstead [16]. In the non-coprime case, Kouvidakis and Pantev [14] had proved a Torelli theorem for $\mathcal{SU}_X(n, L)$, i.e. the corollary under the assumption (1.2), with better bounds. In fact the full corollary can be deduced from this case. However the present line of reasoning is extremely natural, and is of a rather different character from that of Kouvidakis and Pantev. In particular, Theorem 1.0.2 will not follow from their techniques. In the special case where $n = 2$ and $L = \mathcal{O}_X$, Balaji [4] has shown a similar Torelli type theorem for Seshadri's canonical desingularization $N \longrightarrow \mathcal{SU}_X(2, \mathcal{O}_X)$ (see [23]) in the range $g > 3$.[1]

Our strategy in the proof of both theorems is to use a Hecke correspondence to relate the cohomology of $\mathcal{SU}^s_X(n, L)$ with that of another moduli space $\mathcal{SU}_X(n, L'')$ where the degree of L'' is coprime to n. When $n > 2$ the maps defining the Hecke correspondence are only rational. And this necessitates some rather long calculations to bound the codimensions of the indeterminacy loci. Once the basic geometric properties of the correspondence are established, the first theorem follows from some standard arguments in Hodge theory. For the second theorem, we need to make the isomorphism on third cohomology canonical, and to moreover impose an intrinsic polarization on the Hodge structure $H^3(\mathcal{SU}^s_X(n, L))$.

2. The main ideas

For the rest of the paper, we fix a curve X of genus g, $n \in \mathbb{N}$, $d \in \mathbb{Z}$ and a line bundle L of degree d on X. Let $\mathcal{S} = \mathcal{SU}_X(n, L)$ and $\mathcal{S}^s = \mathcal{SU}^s_X(n, L)$.

The main theorems will be proved in the final section of this paper. The broad strategy of our proofs are as follows:

Step 1. *Case* 1. Assume, that d is not divisible by n. Since $\mathcal{SU}_X(n, L)$ is canonically isomorphic to $\mathcal{SU}_X(n, L \otimes \xi^n)$ for every line bundle ξ on X, we may assume that $0 < d < n$.

Fix a set $\chi = \{x^1, \ldots, x^{d-1}\} \subset X$ of $d - 1$ distinct points. Let

$$\mathcal{S}_1 = \mathcal{SU}_X(n, L \otimes \mathcal{O}_X(-D)),$$

where D is the divisor $x^1 + \cdots + x^{d-1}$.

Construct (in § 3) a generalized Hecke correspondence consisting of a pair of rational maps

$$\mathcal{S}_1 \xleftarrow{\pi} \mathbb{P} \overset{\phi}{\dashrightarrow} \mathcal{S}^s. \tag{2.1}$$

By construction, there will be an open subset $U \subset \mathbb{P}$ such that both $\pi|_U$ and $\phi|_U$ will be fiber bundles with fiber isomorphic to $(\mathbb{P}^{n-1})^{d-1}$. Estimates on the codimensions of the complements of U and its image in $\mathcal{S}^s$ will be given in § 4. These estimates, together with some generalities on cohomology and Hodge theory to be established in § 6, will imply

[1] Balaji states the result for $g \geq 3$, but his proof seems to work only for $g > 3$ (see Remark 6.1.1).

that for small i, there are noncanonical isomorphisms of mixed Hodge structures

$$H^i(\mathcal{S}_1,\mathbb{Q}) \cong H^i(\mathcal{S}^s,\mathbb{Q}). \tag{2.2}$$

Therefore this reduces the proof of Theorem 1.0.1 to the case where L and L' have degree 1, and this will be treated in § 7. Moreover, for sufficiently large g, we have isomorphisms modulo torsion of (integral, pure) Hodge structures

$$H^1(X, \mathbb{Z})(-1) \xrightarrow{\sim} H^3(\mathcal{S}_1, \mathbb{Z}) \xrightarrow{\sim} H^3(\mathcal{S}^s, \mathbb{Z}), \tag{2.3}$$

where the first isomorphism is the slant product with a certain universal class (§ 7), and the second now depends canonically on X, L and χ (§ 8). "(-1)" above is the Tate twist. An isomorphism *modulo torsion* of integral pure Hodge structures means that the underlying map of the finitely generated abelian groups is an isomorphism on the free parts. In particular, if as above these Hodge structure have odd weights, the resulting map of the corresponding intermediate Jacobians is an isomorphism.

Case 2. If d is divisible by n then we may assume that $d = 0$. In this case, set $\chi = \{x\}$ for some point $x \in X$. Setting $D = x$, we construct

$$\mathcal{S}_1 = \mathcal{SU}_X(n, L \otimes \mathcal{O}_X(-D))$$

as before. In § 5, we construct a Hecke correspondence analogous to the one above. However, now the map ϕ is regular and the codimension estimates are substantially better. This allows us to establish 2.3 with a much better bound on the genus.

Step 2. Assume that g is chosen sufficiently large. Our first task is to find a (possibly nonprincipal) polarization $\Theta(\mathcal{S}^s)$ on $J^2(\mathcal{S}^s)$ or equivalently on the Hodge structure $H^3(\mathcal{S}^s)$ which varies algebraically with X. The basic tools for constructing this are given in § 6. Let

$$\psi_{X,L,\chi}: H^1(X)(-1) \xrightarrow{\sim} H^3(\mathcal{S}^s)$$

be the isomorphism given above, and

$$\phi_{X,L,\chi} : J(X) \to J^2(\mathcal{S}^s)$$

the corresponding isomorphism of abelian varieties. The isomorphisms vary algebraically with the data (X, L, χ). One can pull $\Theta(\mathcal{S}^s)$ back to get a second polarization on $J(X)$ which varies algebraically with (X, L, χ). If we can find a positive integer m (independent of $(X, L, \chi))$ so that $\Theta(\mathcal{S}^s) = m\Theta$ where Θ is the standard polarization, then it will follow that, after replacing $\Theta(\mathcal{S}^s)$ with $(1/m)\Theta(\mathcal{S}^s)$, that $\Theta(\mathcal{S}^s)$ is a principal polarization such that $(J(X), \Theta) \cong (J^2(\mathcal{S}^s), \Theta(\mathcal{S}^s))$ as required. Since everything varies well, we can assume that X is a sufficiently general curve in moduli. In this case, one checks that any polarization on $J(X)$ is a multiple of the Θ. The precise argument is given in § 8.

3. The biregular Hecke correspondence

The results of this section will be used to treat the case where $d = \deg L$ is not divisible by n. As explained earlier, we may assume that $0 < d < n$. We will continue the notation from step 1 of the previous section. The degree of $L \otimes \mathcal{O}_X(-D)$ is 1, therefore $\mathcal{S}_1$ is smooth and there exists a Poincaré bundle $\mathcal{W}$ on $X \times \mathcal{S}_1$. Let $\mathcal{W}_1, \ldots, \mathcal{W}_{d-1}$ be the $d - 1$

vector bundles on $\mathcal{S}_1$ obtained by restricting $\mathcal{W}$ to $\{x^1\} \times \mathcal{S}_1 = \mathcal{S}_1, \dots, \{x^{d-1}\} \times \mathcal{S}_1 = \mathcal{S}_1$ respectively. Let $\mathbb{P}_k = \mathbb{P}(\mathcal{W}_k)$, $k = 1, \dots, d-1$, where we use the convention $\mathbb{P}(W_i) = \mathrm{Proj}(S^*(W_i^*))$. Let $\mathbb{P}$ $(= \mathbb{P}_{X,L,\chi})$ be the product $\mathbb{P}_1 \times_{\mathcal{S}_1} \dots \times_{\mathcal{S}_1} \mathbb{P}_{d-1}$.

3.1 *The map* $\phi\colon \mathbb{P} \to \mathcal{S}$

We need some notation:

- $\pi\colon \mathbb{P} \to \mathcal{S}_1$ is the natural projection;
- For $1 \le k \le d-1$, $\pi_k\colon \mathbb{P} \to \mathbb{P}_k$ is the natural projection;
- $\imath\colon Z \hookrightarrow X$ is the reduced subscheme defined by $\chi = \{x^1, \dots, x^{d-1}\}$.
- $\imath_k\colon Z_k \hookrightarrow X$, the reduced scheme defined by $\{x_k\}$, $k = 1, \dots, d-1$.
- For any scheme S,

  (i) $p_S\colon X \times S \to S$ and $q_S\colon X \times S \to X$ are the natural projections;
  (ii) $Z^S = q_S^{-1}(Z)$;
  (iii) $Z_k^S = q_S^{-1}(Z_k)$, $k = 1, \dots, d-1$. Note that Z_k^S can be identified canonically with S.

We will show in 3.2 that there is an exact sequence

$$0 \longrightarrow (1 \times \pi)^* \mathcal{W} \longrightarrow \mathcal{V} \longrightarrow \mathcal{T}_0 \longrightarrow 0 \tag{3.1}$$

on $X \times \mathbb{P}$, with $\mathcal{V}$ a vector bundle on $X \times \mathbb{P}$ and $\mathcal{T}_0$ (the direct image of) a line bundle *on the closed subscheme* $Z^{\mathbb{P}}$, which is universal in the following sense: If $\psi\colon S \to \mathcal{S}_1$ is a $\mathcal{S}_1$-scheme and we have an exact sequence

$$0 \longrightarrow (1 \times \psi)^* \mathcal{W} \longrightarrow \mathcal{E} \longrightarrow \mathcal{T} \longrightarrow 0 \tag{3.2}$$

on $X \times S$, with $\mathcal{E}$ a vector bundle on $X \times S$ and $\mathcal{T}$ a line bundle *on the closed subscheme* Z^S, then there is a unique map of $\mathcal{S}_1$-schemes

$$g\colon S \longrightarrow \mathbb{P}$$

such that,

$$(1 \times g)^*(3.1) \equiv (3.2).$$

The $\equiv$ sign above means that the two exact sequences are isomorphic, and the left most isomorphism $(1 \times g)^* \circ (1 \times \pi)^* \xrightarrow{\sim} (1 \times \psi)^*$ is the canonical one.

Let $U_1 \subset \mathbb{P}$ be the maximal open subset such that $\mathcal{V}|_{X \times \{t\}}$ is stable for each $t \in U_1$. We shall see that this is nonempty, thus the natural moduli map $U_1 \to \mathcal{S}^s$ determines a rational map $\phi : \mathbb{P} \to \mathcal{S}^s$. The geometric properties of the map $\phi|_{U_1}$ are not obvious, so in 3.3 we will shrink U_1 to an open subset U with more manageable properties.

3.2 *The universal exact sequence*

We begin by reminding the reader of some elementary facts from commutative algebra. If A is a ring (commutative, with 1), $t \in A$ a non-zero divisor, and M an A-module, then each element $m_0 \in M$ gives rise to an equivalence class of extensions

$$0 \longrightarrow M \longrightarrow E_{m_0} \longrightarrow A/tA \longrightarrow 0, \tag{3.3}$$

where $E_{m_0} = (A \oplus M)/A(t, m_0)$, and the arrows are the obvious ones. Moreover, if $m_0 - m_1 \in tM$, say

$$m_0 - m_1 = tm',$$

then the extension given by m_0 is equivalent to that given by m_1. In fact, one checks that

$$\begin{aligned} &E_{m_0} \xrightarrow{\sim} E_{m_1} \\ &(a,m) \mapsto (a, m - am') \end{aligned} \tag{3.4}$$

gives the desired equivalence of extensions. This is another way of expressing the well-known fact that each element of $M/tM = \mathrm{Ext}^1(A/t, M)$ gives rise to an extension.

One globalizes to get the following : Let S be a scheme, $T \overset{\imath}{\hookrightarrow} S$ a closed immersion, $\mathcal{F}$ a quasi-coherent $\mathcal{O}_S$-module, U an open neighbourhood of T in S, and $t \in \Gamma(U, \mathcal{O}_S)$ an element which defines $T \hookrightarrow U$, and which is a non-zero divisor for $\Gamma(V, \mathcal{O}_S)$ for any open $V \subset U$. Then every global section s of $\imath^*\mathcal{F} = \mathcal{F} \otimes \mathcal{O}_T$ gives rise to an equivalence class of extensions

$$0 \longrightarrow \mathcal{F} \longrightarrow \mathcal{E} \longrightarrow \mathcal{O}_T \longrightarrow 0. \tag{3.5}$$

Indeed, we are reduced immediately to the case $S = U$. We build up exact sequences (3.3) on each affine open subset $W \subset S$, by picking a lift $\tilde{s_W} \in \Gamma(W, \mathcal{F})$ of $s \mid W$. One patches together these exact sequences via (3.4).

Now consider $\mathbb{P} = \mathbb{P}_1 \times_{\mathcal{S}_1} \cdots \times_{\mathcal{S}_1} \mathbb{P}_{d-1}$. For each $k = 1, \ldots, d-1$, let $p_k : \mathbb{P}_k \to \mathcal{S}_1$ be the natural projection. We have a universal exact sequence

$$0 \longrightarrow \mathcal{O}(-1) \longrightarrow p_k^* \mathcal{W}_k \longrightarrow B \longrightarrow 0$$

whence a global section $s_k \in \Gamma(\mathbb{P}_k, p_k^* \mathcal{W}_k(1))$. However, note that

$$p_k^* \mathcal{W}_k = (1 \times p_k)^* \mathcal{W} \mid Z_k^{\mathbb{P}_k},$$

where we are identifying $Z_k^{\mathbb{P}_k}$ with $\mathbb{P}_k$. By (3.5) we get exact sequences

$$0 \longrightarrow (1 \times \pi)^* \mathcal{W} \longrightarrow \mathcal{V}_k \longrightarrow \mathcal{O}_{Z_k^{\mathbb{P}}} \otimes L_k \longrightarrow 0,$$

where L_k is the line bundle obtained by pulling up $\mathcal{O}_{\mathbb{P}_k}(-1)$. It is not hard to see that $\mathcal{V}_k$ is a family of vector bundles parameterized by $\mathbb{P}$. Gluing these sequences together — the kth and the lth agree outside $Z_k^{\mathbb{P}}$ and $Z_l^{\mathbb{P}}$ — we obtain (3.1).

Now suppose we have a $\mathcal{S}_1$-scheme $\psi : S \to \mathcal{S}_1$ and the exact sequence (3.2)

$$0 \longrightarrow (1 \times \psi)^* \mathcal{W} \longrightarrow \mathcal{E} \longrightarrow \mathcal{T} \longrightarrow 0$$

on $X \times S$. Restricting (3.2) to Z_k^S $(1 \le k \le d-1)$ one checks that the kernel of $(1 \times \psi)^* \mathcal{W} \mid Z_k^S \to \mathcal{E} \mid Z_k^S$ is a line bundle $\mathcal{L}_k$. Identifying Z_k^S with S, we see that $(1 \times \psi)^* \mathcal{W} \mid Z_k^S = \psi^* \mathcal{W}_k$. Thus $\mathcal{L}_k$ is a line sub-bundle of $\psi^* \mathcal{W}_k$. By the universal property of $\mathbb{P}_k$, we see that we have a unique map of $\mathcal{S}_1$-schemes

$$g_k : S \longrightarrow \mathbb{P}_k$$

such that $\mathcal{O}(-1)$ gets pulled back to $\mathcal{L}_k$. The various g_k give us a map

$$g : S \longrightarrow \mathbb{P}.$$

One checks that g has the required universal property. The uniqueness of g follows from the uniqueness of each g_k.

3.3 *The biregular Hecke correspondence*

As explained earlier, we will need to consider an open subset $U \subset U_1$, with good geometric properties. Essentially U will be the preimage of a subset $U' \subset \mathcal{S}^s$ parameterizing those stable vector bundles E for which the kernel of $p : E \to \mathcal{O}_Z$ is stable for every surjection. We again remind the reader, that if $\deg L$ is a multiple of n, one can do better, as will be seen in § 5.

To construct U and U' rigorously, it would be convenient to have some sort of universal family of vector bundles on $X \times \mathcal{S}^s$. Unfortunately, such a family will not exist when $(n, d) \neq 1$. To get around this, we can go back to the construction ([22]): $\mathcal{S}^s$ is obtained as a good quotient of an open subscheme R of a Quot scheme by a reductive group G (which is in fact a projective general linear group). There is a vector bundle $\mathbb{E}'$ on $X \times R$ whose restriction to $X \times \{r\}$ is the bundle represented by the image of r in $\mathcal{S}^s$. For $x \in X$, let E_x be the restriction of $\mathbb{E}'$ to $R \cong \{x\} \times R$. There is a natural G action on $\mathbb{E}'$ which induces one on E_x for every $x \in X$ and hence on every $\mathbb{P}(E_x^*)$. This action is locally a product of the action on R with a trivial action along the fibers, therefore $\mathbb{P}(E_x^*)/G$ is a projective space bundle (or more accurately a Brauer–Severi scheme) over $\mathcal{S}^s$. Let $\pi' : \mathbb{P}' \to R$ be the fiber product of $\mathbb{P}(E_{x^i}^*) \to R$ for $i = 1, \ldots, d-1$. Then $\mathbb{P}'/G \to \mathcal{S}^s$ is a $(\mathbb{P}^{n-1})^{d-1}$-bundle i.e. a smooth morphism with fibers isomorphic to $(\mathbb{P}^{n-1})^{d-1}$. Let L_i be the pullback of $\mathcal{O}_{\mathbb{P}(E_{x^i}^*)}(1)$ to $\mathbb{P}'$. There is a canonical map

$$\kappa : (1 \times \pi')^* \mathbb{E}' \to \bigoplus_k i_{k*} L_k,$$

where i_k: $\mathbb{P}' = \{x^k\} \times \mathbb{P}' \to X \times \mathbb{P}'$ is the natural closed immersion.

Let $\mathcal{U}''$ be the maximal open subset of $\mathbb{P}'$ such that $\ker(\kappa)|_{X \times \{t\}}$ is stable for each $t \in \mathcal{U}''$. Set $\mathcal{U}' = R \setminus \pi'(\mathbb{P}' \setminus \mathcal{U}'')$ and $\mathcal{U} = \pi'^{-1}\mathcal{U}' \subset \mathcal{U}''$. Both $\mathcal{U}$ and $\mathcal{U}'$ are invariant under G, and we let U and $U' \subset \mathcal{S}^s$ be the quotients. The map $U \to U'$ is again a $(\mathbb{P}^{n-1})^{d-1}$-bundle. The exact sequence

$$0 \to \ker(\kappa) \to (1 \times \pi')^* \mathbb{E}' \to \bigoplus_k i_{k*} L_k \to 0$$

yields, via the universal property of $\mathbb{P}$ a morphism $\mathcal{U} \to \mathbb{P}$ which factors through U. The map $U \to \mathbb{P}$ is an open immersion to a subset of U_1. Thus we get a diagram

$$\mathcal{S}_1 \leftarrow U \xrightarrow{f} \mathcal{S}^s \tag{3.6}$$

which will be referred to as the *biregular Hecke correspondence.*

Remark 3.3.1. With the above notation, note that $\mathbb{P}'/G \to \mathcal{S}^s$ is the fibre product of $\mathbb{P}(E_{x^i}^*)/G \to \mathcal{S}^s (i = 1, \ldots, d-1)$. It follows that f itself may be regarded as the fibre product of $\mathbb{P}^{n-1}$ bundles over U'.

4. Codimension estimates

We will continue the notation from the previous section. Our goal is to establish the basic estimates on the codimensions of the complements of U and U'. In particular, these estimates will show that U and U' are nonempty.

4.1 *Special subvarieties*

Let X be a smooth curve of genus $g \geq 2$. Fix integers $n > m \geq 1$, a rational number μ_0 and a line bundle M. Let $T_{m,\mu_0}(n, M)$ be the subset of $\mathcal{SU}^s(n, M)$ whose points correspond to bundles E for which there exists a subbundle $F \subset E$ of rank m and slope $\mu(F) = \mu_0$.

Lemma 4.1.1. *$T_{m,\mu_0}(n,M)$ is Zariski closed, and can therefore be regarded as a reduced subscheme. There exists a scheme Σ such that $X \times \Sigma$ carries a rank n vector bundle $\mathbb{E}$ with a rank m subbundle $\mathbb{F}$. For each $\sigma \in \Sigma$, the restriction of $\mathbb{E}$ to $X \times \{\sigma\}$ is stable of degree d, while the restriction of $\mathbb{F}$ has slope μ_0. The canonical map $\Sigma \to \mathcal{SU}^s(n,M)$ has T_{m,μ_0} as its image.*

Proof. The argument is similar to that used above. $S = \mathcal{SU}^s(n,M)$ can be realized as a good quotient of a subscheme R of a Quot scheme by a reductive group G. $X \times R$ carries a vector bundle $\mathbb{E}'$ whose restriction to $X \times \{r\}$ is the bundle represented by the image of r in S. $\mathbb{E}'$ extends to a coherent sheaf (denoted by the same symbol) on the closure $Y = X \times \bar{R}$. Let Q be the closed subscheme of the relative Quot scheme $\mathrm{Quot}_{X \times \bar{R}/\bar{R}}(\mathbb{E}')$ parameterizing quotients which restrict to vector bundles of degree $\deg M - \mu_0 m$ and rank $n-m$ on each $X \times \{r\}$. The intersection $\tilde{S}$ of R with the image of the projection $Q \to \bar{R}$ is closed and G-invariant. Therefore the image of $\tilde{S}$ in S, which is T_{m,μ_0}, is closed.

Set $\Sigma = R \times_{\bar{R}} Q$ and $\mathbb{E}$ to the pullback of $\mathbb{E}'$ to $X \times \Sigma = (X \times R) \times_{\bar{R}} Q$. Then it is easily seen that these have the required properties. □

4.2 *Deformations*

Let $D = \mathrm{Spec}\, \mathbb{C}[\epsilon]/(\epsilon^2)$, and let $\mathcal{E}$ be a vector bundle on $X \times D$. X can be identified with a closed subscheme of $X \times D$ with ideal sheaf $\mathcal{I} = \epsilon \mathcal{O}_{X \times D}$. Let $E = \mathcal{E} \otimes \mathcal{O}_X \cong \mathcal{E} \otimes \mathcal{I}$. Then there is an exact sequence

$$0 \to E \to \mathcal{E} \to E \to 0$$

which is classified by an element of $\mathrm{Ext}^1(E,E) \cong H^1(\mathcal{E}nd(E))$. Furthermore

$$0 \to \det(E) \to \det(\mathcal{E}) \to \det(E) \to 0$$

is classified by the trace of the above class. If $\mathcal{E}nd_0(E)$ denotes the traceless part of $\mathcal{E}nd(E)$, then $H^1(\mathcal{E}nd_0(E))$ classifies deformations of E which induce trivial deformations of $\det(E)$.

If $M = \det E$, then elements of $H^1(X, \mathcal{E}nd_0(E))$ give rise to maps from D to $\mathcal{SU}(n,M)$ which send the closed point to the class of E. This map is well known to yield an isomorphism between $H^1(X, \mathcal{E}nd_0(E))$ and the tangent space to $\mathcal{SU}(n,M)$ at $[E]$.

Let E be a vector bundle corresponding to general point $[E]$ of a component of T_{m,μ_0}, and let v be a tangent vector to T_{m,μ_0} based at $[E]$. This vector can be lifted to tangent vector $\tilde{v}$ to Σ at a point σ lying over $[E]$. Then $E \cong \mathbb{E}|_{X \times \{\sigma\}}$, and let F be the subbundle corresponding to $\mathbb{F}|_{X \times \{\sigma\}}$. Set $G = E/F$,

$$K = \ker[\mathcal{E}nd(E) \to \mathcal{H}om(F,G)]$$

and

$$K_0 = K \cap \mathcal{E}nd_0(E) = \ker[\mathcal{E}nd_0(E) \to \mathcal{H}om(F,G)].$$

Note that $K = K_0 \oplus (id) \cdot \mathcal{O}_X$.

Let $\mathcal{E}$ be the first order deformation of E corresponding to the tangent vector v. Explicitly, if $D \to \Sigma$ is the map corresponding to $\tilde{v}$, then $\mathcal{E} = \mathbb{E}|_D$. Setting $\mathcal{F} = \mathbb{F}|_D$ gives a deformation of F which fits into a diagram:

$$\begin{tikzcd}
0 \arrow[r] & E \arrow[r] & \mathcal{E} \arrow[r] & E \arrow[r] & 0 \\
0 \arrow[r] & F \arrow[r] \arrow[u] & \mathcal{F} \arrow[r] \arrow[u] & F \arrow[r] \arrow[u] & 0.
\end{tikzcd}$$

The images of the classes of $\mathcal{F}$ and $\mathcal{E}$ in $\mathrm{Ext}^1(F,E)$ agree up to sign. Therefore the class of $\mathcal{E}$ lies in the kernel of the map to $\mathrm{Ext}^1(F,G)$ which is the image of $H^1(X,K_0)$.

Lemma 4.2.1. *The image of $H^1(X,K_0)$ in $H^1(X,\mathcal{E}nd_0(E))$ contains the tangent space to T_{m,μ_0} at E.*

A simple diagram chase shows that there is an exact sequence

$$0 \to \mathcal{H}om(E,F) \to K \to \mathcal{E}nd(G) \to 0.$$

COROLLARY 4.2.1

$\dim T_{m,\mu_0} \leq h^1(\mathcal{E}nd(F)) + h^1(\mathcal{H}om(G,F)) + h^1(\mathcal{E}nd(G)) - g.$

4.3 *Preliminary codimension estimates*

Let us say that T_{m,μ_0} is admissible if for any general point $[E]$ of a component of T_{m,μ_0} of largest dimension, there exists a semistable subbundle $F \subset E$ of rank m and slope μ_0. We will estimate codimension of an admissible $T_{m,\mu_0}(n,M)$ from below as a function of four quantities $g \geq 2$, $n \geq 2$, $n > m \geq 1$, and $\mu_0 < \mu = n/\deg M$. The imposition of admissibility simplifies the calculations, and presents no real loss of generality. Fixing E and F as above, let $G = E/F$ and let $0 = G_0 \subset G_1 \subset \cdots \subset G_r = G$ be the Harder–Narasimhan filtration. Set $n_i = \mathrm{rk}(G_i/G_{i-1})$ and $n_0 = m$. We have

$$\mu < \mu(G_r/G_{r-1}) < \cdots < \mu(G_1) < \frac{n_0}{n_1}(\mu - \mu_0) + \mu,$$

where the last inequality follows from the bound on the slope of the preimage of G_1 in E.

In the computations below, we will make use of the additivity of 'deg' and 'rk'.

Lemma 4.3.1. *If A and B are locally free then $\mu(A \otimes B) = \mu(A) + \mu(B)$ and $\mu(\mathcal{H}om(A,B)) = \mu(B) - \mu(A)$.*

Lemma 4.3.2. *If V is semi-stable, then $h^0(V) \leq \deg(V) + \mathrm{rk}(V)$ provided that the right side is nonnegative.*

Proof. This is obvious if $\deg V < 0$, since $h^0(V) = 0$. For $\deg V \geq 0$, we can assume that the lemma holds for $V(-p)$ by induction. Therefore

$$h^0(V) \leq h^0(V(-p)) + h^0(V/V(-p)) \leq (\deg V - r + r) + r.$$

□

COROLLARY 4.3.1

If V is a semistable vector bundle then $h^1(\mathcal{E}nd(V)) \leq \mathrm{rk}(V)^2 g$.

Proof. As $\mathcal{E}nd(V)$ is semistable, we have $h^0(\mathcal{E}nd(V)) \leq \mathrm{rk}(V)^2$, and the result now follows from Riemann–Roch. □

Heuristically, $\dim T_{m,\mu_0}$ should be given by the number of moduli for F, G_i plus extensions. To make this more rigorous, we will work infinitesimally, and appeal to

Corollary 4.2.1. Each of the terms of the corollary can be estimated in turn. For the first term, we have

$$h^1(\mathcal{E}nd(F)) \leq n_0^2 g. \tag{4.1}$$

$\operatorname{Hom}(G,F) = 0$ by the numerical conditions, therefore

$$h^1(\mathcal{H}om(G,F)) = -\chi(G^* \otimes F).$$

So Riemann–Roch yields

$$h^1(\mathcal{H}om(G,F)) = n_0(n_1 + \cdots + n_r)(g-1) + nn_0(\mu - \mu_0). \tag{4.2}$$

Note deg $\mathcal{H}om(G,F) =$ deg $\mathcal{H}om(E,F) = -nn_0(\mu - \mu_0)$.

The last term $h^1(\mathcal{E}nd(G))$ remains. If $G = G_1$ is semi-stable, then a bound is given as above. To analyse the general case, we need the following lemma.

Lemma 4.3.3. *If V has a filtration such that the associated graded sheaves are semi-stable bundle with slope at least -1, then $h^1(V) \leq g \cdot \operatorname{rk}(V)$.*

Proof. By subadditivity of h^1, it is enough to assume that V is semi-stable with slope at least -1, in which case the result follows from the previous lemma together with Riemann–Roch. □

COROLLARY 4.3.2

Let W be a semi-stable bundle, and V a vector bundle with a filtration such that the associated graded sheaves are semi-stable bundle with slope at least $\mu(W) - 1$, then

$$h^1(\mathcal{H}om(W,V)) \leq \operatorname{rk}(V)\operatorname{rk}(W)g.$$

Lemma 4.3.4. *We have the inequality*

$$h^1(\mathcal{E}nd(G_k)) \leq (n_1 + \cdots + n_k)^2 g + \left(\sum_{1\leq i<j\leq k} n_i n_j\right)(\mu(G_1) - \mu(G) - 1).$$

Proof. From

$$0 \to G_{k-1} \to G_k \to G_k/G_{k-1} \to 0,$$

we obtain

$$\begin{aligned} h^1(\mathcal{E}nd(G_k)) \leq\, & h^1(\mathcal{E}nd(G_{k-1})) + h^1(\mathcal{E}nd(G_k/G_{k-1})) \\ & + h^1(\mathcal{H}om(G_{k-1}, G_k/G_{k-1})) + h^1(\mathcal{H}om(G_k/G_{k-1}, G_{k-1})). \end{aligned}$$

It suffices to find upper bounds for each of the terms on the right, and then sum them. The first term can be controlled by induction, and the second and third terms by the previous corollaries. For the last term, an estimate can be obtained by combining $\operatorname{Hom}(G_k/G_{k-1}, G_{k-1}) = 0$, with

$$-\deg \mathcal{H}om(G_k/G_k - 1, G_k - 1) = (n_1 + \cdots + n_{k-1})n_k[\mu(G_k/G_{k-1}) - \mu(G_{k-1})]$$

$$
\begin{aligned}
&= \sum_{i=1}^{k-1} n_i n_k [\mu(G_k/G_{k-1}) - \mu(G_i/G_{i-1})] \\
&< \sum_{i=1}^{k-1} n_i n_k [\mu(G_1) - \mu]
\end{aligned}
$$

and the Riemann–Roch theorem. □

Thus

$$
h^1(\mathcal{E}nd(G_r)) - g \le \left[(n_1 + \cdots + n_r)^2 - 1\right] g + \left(\sum_{1 \le i < j \le r} n_i n_j \right) (\mu(G_1) - \mu - 1). \tag{4.3}
$$

Subtracting eqs (4.1), (4.2) and (4.3) from dim $\mathcal{SU}(n, M)$ and simplifying, and replacing n_0 by m, yields

$$
\begin{aligned}
\operatorname{codim} T_{m,\mu_0} > m(n-m)(g+1) - \left(\sum_{1 \le i < j \le r} n_i n_j \right) (\mu(G_1) - \mu(G) - 1) \\
- nm(\mu - \mu_0) - (n^2 - 1).
\end{aligned}
$$

To proceed further, we use $\sum_{1 \le i < j \le r} n_i n_j \le \frac{1}{2}(n-m)^2$ and

$$
\mu(G_1) - \mu \le \frac{m}{n_1}(\mu - \mu_0) \le m(\mu - \mu_0).
$$

Putting these together yields the following.

PROPOSITION 4.3.1

If T_{m,μ_0} is admissible, we have

$$
\operatorname{codim} T_{m,\mu_0} > m(n-m)(g+1) - \tfrac{1}{2} m(n-m)^2 (\mu - \mu_0) - nm(\mu - \mu_0) - (n^2 - 1).
$$

4.4 *Final codimension estimates*

Recall that in § 3 we had a diagram

$$
\mathcal{S}_1 \xleftarrow{\pi} \mathbb{P} \supset U_1 \supset U \xrightarrow{f} U' \subset \mathcal{S}^s,
$$

where $\mathbb{P}$ could be viewed as the parameter space for extensions

$$
0 \to E \to E' \to \mathcal{O}_Z \to 0
$$

with $E \in \mathcal{S}_1$. An extension E' lies in $\mathbb{P} \setminus U_1$ if and only if there exists a subbundle $F' \subset E'$ with $\mu(F') \ge \mu(E') = \mu(E) + (d-1)/n$. F' can be assumed to be semi-stable, since otherwise it can be replaced by the first step in the Harder–Narasimhan filtration. Let $F = F' \cap E$, then

$$
\mu(F) \ge \mu(F') - \frac{(d-1)}{m} \ge \mu(E) - (d-1)\left(\frac{1}{m} - \frac{1}{n}\right) \ge \mu(E) - \frac{n-1}{m},
$$

where $m = \operatorname{rk}(F)$.

Similarly, if an extension E' lies in $U_1 \setminus U$ then there exist a coherent subsheaf $F_1' \subset E'$ which violates stability of the kernel of some map $E' \to \mathcal{O}_Z$. After replacing F_1' by the maximal semistable subbundle of its saturation, and setting $F_1 = F_1' \cap E$, we obtain

$$\mu(F_1) \geq \mu(F_1') - \frac{(d-1)}{m'} \geq \mu(E) - \frac{(d-1)}{m'} \geq \mu(E) - \frac{n-1}{m'},$$

where $m' = \mathrm{rk}(F_1)$.

Therefore $\pi(\mathbb{P} \setminus U)$ is contained in the union of admissible $T_{m,\mu_0}(n, L \otimes \mathcal{O}_X(-D))$, as m varies between 1 and $n-1$, and $\mu - \mu_0$ between zero and $(n-1)/m$. Clearly $\mathrm{codim}(\mathbb{P} \setminus U) \geq \mathrm{codim}(\pi(\mathbb{P} \setminus U))$. Therefore a term by term estimate of the bound in Proposition 4.3.1, using $(n-m) \leq n-1$ and $m(n-m) \geq n-1$ (for $1 \leq m \leq n-1$), yields:

Theorem 4.4.1. $\mathrm{codim}(\mathbb{P} \setminus U) > (n-1)[g - \frac{1}{2}(n^2+n)]$.

COROLLARY 4.4.1

If

$$g > \frac{k}{n-1} + \frac{n^2+n}{2}$$

then,

$$\mathrm{codim}(\mathbb{P} \setminus U) > k.$$

The estimate on the codimension of U' is obtained by a similar argument. If $E' \in \mathcal{S} \setminus U'$, then there exists a surjection $E' \to \mathcal{O}_Z$ and a subbundle $F \subset \ker[E' \to \mathcal{O}_Z]$ violates the stability of the kernel. Let $m = \mathrm{rk}(F)$ and F' be the maximal semistable subbundle of the saturation of F in E', then

$$\begin{aligned}
\mu(F) &\geq \mu(F') - \frac{(d-1)}{m} \\
&\geq \mu(E) - \frac{(d-1)}{m} \\
&= \mu(E') - (d-1)\left(\frac{1}{n} + \frac{1}{m}\right) \\
&\geq \mu(E') - (n-1)\left(\frac{1}{n} + \frac{1}{m}\right).
\end{aligned}$$

Thus the complement of U' lies in the union of admissible $T_{m,\mu_0}(n, L)$ as m varies between 1 and $n-1$, and $\mu - \mu_0$ varies between zero and $(n-1)(\frac{1}{n} + \frac{1}{m})$. An elementary analysis shows that

$$\left(1 + \frac{m}{n}\right)\left[\frac{1}{2}(n-m)^2 + n\right] \leq \left(1 + \frac{1}{n}\right)\left[\frac{1}{2}(n-1)^2 + n\right]$$

for $1 \leq m \leq n-1$ and $n \geq 4$. Combining this and the trivial estimate $m(n-m) \geq n-1$ with Proposition 4.3.1 yields:

Theorem 4.4.2. *If $n \geq 4$ then*

$$\operatorname{codim}(\mathcal{S}^s \backslash U) > (n-1)\left[g - \frac{n^2+3n+1}{2} - \frac{3}{n}\right] > (n-1)\left[g - \frac{n^2+3n+1}{2}\right] - 3.$$

COROLLARY 4.4.2

If $n \geq 4$ and $g > (k/(n-1)) + ((n^2+3n+3)/2)$, then $\operatorname{codim}(\mathcal{S}^s \backslash U') > k$.

Of course, the above inequality for g implies the inequality in Corollary 4.4.1. The restriction on n above is harmless, because the cases of $n = 2, 3$ can be handled by the results of the next section.

5. Hecke when $\deg L = 0$

The notations in this section are special to this section. We now give an improved Hecke correspondence when $\deg L \in n\mathbb{Z}$. Clearly, without loss of generality, we may (and will) assume that $\deg L = 0$. In this case, instead of choosing $d-1$ points on X (which clearly does not make sense), we choose one point $x \in X$, and let Z be the reduced scheme supported on $\{x\}$. Let D be the divisor given by $\{x\}$. Then $\deg L \otimes \mathcal{O}_X(-D)$ is -1, and if $\mathcal{S}_1 = \mathcal{SU}_X(n, L \otimes \mathcal{O}_X(-D))$, then $\mathcal{S}_1$ is smooth and there exists a Poincaré bundle $\mathcal{W}$ on $X \times \mathcal{S}_1$. Let $\mathcal{W}_1$ be the vector bundle on $\mathcal{S}_1$ obtained by restricting $\mathcal{W}$ to $\{x\} \times \mathcal{S}_1 = \mathcal{S}_1$, and $\mathbb{P} = \mathbb{P}(\mathcal{W}_1)$. Let $\pi : \mathbb{P} \to \mathcal{S}_1$ be the natural projection. Then as before we have the universal exact sequence (3.1)

$$0 \to (1 \times \pi)^* \mathcal{W} \to \mathcal{V} \to \mathcal{T}_0 \to 0,$$

of coherent sheaves on $X \times \mathbb{P}$ with $\mathcal{V}$ a vector bundle, and $\mathcal{T}_0$ a sheaf supported on $\{x\} \times \mathbb{P} = \mathbb{P}$ which is a line bundle on $\mathbb{P}$. This means that $\mathbb{P}$ parameterizes exact sequences

$$0 \to W \to V \to \mathcal{O}_Z \to 0$$

of coherent sheaves on X with $W \in \mathcal{S}_1$ and V a vector bundle (necessarily of rank n and determinant L).

There is a way of interpreting this universal property in terms of quasi-parabolic bundles (see [15, p. 211–212, Definition 1.5] for the definitions of quasi-parabolic and parabolic bundles). We introduce a quasi-parabolic datum on X by attaching the flag type $(1, n-1)$ to the point x. From now onwards quasi-parabolic structures will be with respect to this datum and on vector bundles of rank n and determinant L. One observes that for a vector bundle V (of rank n and determinant L), a surjective map $V \twoheadrightarrow \mathcal{O}_Z$ determines a unique quasi-parabolic structure, and two such surjections give the same quasi-parabolic structure if and only if they differ by a scalar multiple. The above mentioned universal property says that $\mathbb{P}$ is a (fine) moduli space for quasi-parabolic bundles. More precisely, the family of quasi-parabolic structures

$$\mathcal{V} \twoheadrightarrow \mathcal{T}_0$$

parameterized by $\mathbb{P}$ is universal for families of quasi-parabolic bundles

$$\mathcal{E} \twoheadrightarrow \mathcal{T}$$

parameterized by S, whose kernel is a family of semi-stable bundles. The points of $\mathbb{P}$ parameterize quasi-parabolic structures $V \twoheadrightarrow \mathcal{O}_Z$ whose kernel is semi-stable.

Let $\alpha = (\alpha_1, \alpha_2)$, where $0 < \alpha_1 < \alpha_2 < 1$, and let $\Delta = \Delta_\alpha$ be the parabolic datum which attaches weights α_1, α_2 to our quasi-parabolic datum. We can choose α_1 and α_2 so small that

- a parabolic semi-stable bundle is parabolic stable;
- if V is stable, then every parabolic structure on V is parabolic stable;
- the underlying vector bundle of a parabolic stable bundle is semi-stable in the usual sense.

Showing the above involves some very elementary calculations. Call α *small* if α_1 and α_2 satisfy the above properties and denote the resulting (fine) moduli space of parabolic stable bundles $\mathcal{SU}_X(n, L, \Delta)$.

Lemma 5.0.1. *If α is small then for every parabolic stable bundle $V \twoheadrightarrow \mathcal{O}_Z$, the kernel W is semi-stable.*

Proof. Note that $\deg W = -1$ and hence the semi-stability of W is equivalent to its stability. Suppose W is not stable. Then by the above observation, there is a subbundle E of W such that $\mu(E) > \mu(W)$. Let $\operatorname{rk} E = r$. Let $E' \subset V$ be the subbundle generated by E. Let T be the torsion subsheaf of the cokernel of $E \to V$, and t the vector space dimension (over $k(x) = \mathbb{C}$) of T. We then have $\deg E' = \deg E + t$, where $t \geq 0$. Thus

$$-\frac{1}{n} < \mu(E) \leq \mu(E') = \mu(E) + \frac{t}{r} \leq \mu(V) = 0.$$

In particular

$$-\frac{1}{n} < \mu(E) \leq -\frac{t}{r},$$

i.e.

$$\frac{1}{n} > \frac{t}{r},$$

but this is not possible for $t > 0$ since $r < n$. So $t = 0$ and hence $T = 0$. Thus $E = E'$. Let $d = \deg E$. Then

$$-\frac{1}{n} < \frac{d}{r} \leq 0.$$

This is possible only if $d = 0$. Since $E = E'$, one checks that the flag induced on the fibre E'_x by the flag $F_1V_x \supset F_2V_x$ is the trivial flag $E'_x = F_1E'_x = F_2E'_x$. It follows that the parabolic degree of E' is $r\alpha_2$. Since V is parabolically stable, this means

$$\alpha_2 < \frac{\alpha_1 + (n-1)\alpha_2}{n}.$$

The right side is a non-trivial convex combination of α_1 and α_2, and we also have $\alpha_1 < \alpha_2$. Thus we have a contradiction. Therefore W is stable (see also [3] for the same result when $n = 2$). □

Theorem 5.0.3. *$\mathbb{P} = \mathcal{SU}_X(n, L, \Delta)$, and $\mathcal{V} \twoheadrightarrow \mathcal{T}_0$ is the universal family of parabolic bundles.*

Proof. Let $\mathbb{P}^{ss} \subset \mathbb{P}$ be the locus on which $\mathcal{V}$ consists of parabolic semi-stable (= parabolic stable) bundles. One checks that $\mathbb{P}^{ss}$ is an open subscheme of $\mathbb{P}$ (this involves two things: (i) knowing that the scheme $\widetilde{R}$ of [15, p. 226] has a local universal property for parabolic bundles and (ii) knowing that the scheme $\widetilde{R}^{ss}$ of *loc. cit.* is open).

Clearly $\mathbb{P}^{ss}$ is non-empty — in fact if V is stable of rank n and determinant L, then any parabolic structure on V is parabolic stable (see above). We claim that $\mathbb{P}^{ss} \simeq \mathcal{SU}_X$ (n, L, Δ). To that end, let S be a scheme, and

$$\mathcal{E} \twoheadrightarrow \mathcal{T} \tag{5.1}$$

a family of parabolic stable bundles parameterized by S. By Lemma 5.0.1, the kernel $\mathcal{W}'$ of (5.1) is a family of stable bundles of rank n and determinant $L \otimes \mathcal{O}_X(-D)$. Since $\mathcal{S}_1$ is a fine moduli space, we have a unique map $g: S \to \mathcal{S}_1$ and a line bundle ξ on S such that $(1 \times g)^* \mathcal{W} = \mathcal{W}' \otimes p_S^* \xi$. By doctoring (5.1) we may assume that $\xi = \mathcal{O}_S$. The universal property of the exact sequence (3.1) on $\mathbb{P}$ then gives us a unique map

$$g: S \longrightarrow \mathbb{P}$$

such that $(1 \times g)^*(3.1)$ is equivalent to

$$0 \longrightarrow \mathcal{W}' \longrightarrow \mathcal{E} \longrightarrow \mathcal{T} \longrightarrow 0.$$

Clearly g factors through $\mathbb{P}^{ss}$. This proves that $\mathbb{P}^{ss}$ is $\mathcal{SU}_X(n, L, \Delta)$. However, $\mathcal{SU}_X(n, L, \Delta)$ is a projective variety (see [15], pp. 225–226, Theorem 4.1), whence we have

$$\mathbb{P} = \mathcal{SU}_X(n, L, \Delta).$$

Clearly, $\mathcal{V} \twoheadrightarrow \mathcal{T}_0$ is the universal family of parabolic bundles. □

Note that the above proof gives $\mathbb{P}^{ss} = \mathbb{P}$.

COROLLARY 5.0.3

Let

$$0 \to W \to V \to \mathcal{O}_Z \to 0$$

be an exact sequence of coherent sheaves on X with $W \in \mathcal{S}_1$ and V a vector bundle. Then V is a semi-stable bundle.

It follows that $\mathcal{V}$ consists of (usual) semi-stable bundles (by our choice of α). Since $\mathcal{S}$ is a coarse moduli space, we get the map

$$\varphi: \mathbb{P} \longrightarrow \mathcal{S}. \tag{5.2}$$

Remark 5.0.1. Note that the parabolic structure Δ is something of a red herring. In fact $\mathcal{SU}_X(n, L, \Delta)$ parameterizes quasi-parabolic structures $V \twoheadrightarrow \mathcal{O}_Z$, whose kernel is semi-stable (cf. [15, p. 238, Remark (5.4)] where this point is made for $n = 2$).

Remark 5.0.2. Let V be a stable bundle of rank n, with $\det V = L$, so that (the isomorphism class of) V lies in $\mathcal{S}^s$. Since any parabolic structure on V is parabolic stable (by our choice of α), therefore we see that $f^{-1}(V)$ is canonically isomorphic to $\mathbb{P}(V_x^*)$.[2] Moreover, it is not hard to see that $\mathbb{P}^s := \pi^{-1}(\mathcal{S}^s) \to \mathcal{S}^s$ is smooth (examine the effect on the tangent space of each point on $\mathbb{P}^s$).

5.1 *Codimension estimates*

We wish to estimate $\operatorname{codim}(\mathbb{P} \setminus \mathbb{P}^s)$ as well as $\operatorname{codim}(\mathcal{S} \setminus \mathcal{S}^s)$. The second admits to exact answers (see Remark 5.1.1 below). Heuristically (one can make this rigorous via the deformation theoretic techniques in § 4), the method for obtaining the first estimate is as follows.

Let $V \twoheadrightarrow \mathcal{O}_Z$ be a parabolic bundle in $\mathbb{P} \setminus \mathbb{P}^s$. Then we have a filtration (see [22], p. 18, Théorème 10)

$$0 = V_{p+1} \subset V_p \subset \cdots \subset V_0 = V$$

such that for $0 \leq i \leq p$, $G_i = V_i/V_{i+1}$ is stable and $\mu(G_i) = \mu$. Moreover (the isomorphism class of) the vector bundle $\oplus\, G_i$ depends only upon V and not on the given filtration. We wish to count the number of moduli at $[V \overset{\theta}{\twoheadrightarrow} \mathcal{O}_Z] \in \mathbb{P} \setminus \mathbb{P}^s$. There are three sources:

(a) The choice of $\oplus_{i=0}^{p} G_i$;
(b) Extension data;
(c) The choice of parabolic structure $V \overset{\theta}{\twoheadrightarrow} \mathcal{O}_Z$, for fixed semi-stable V.

The source (c) is the easiest to calculate — there is a codimension one subspace at each parabolic vertex, contributing $n-1$ moduli. Let $n_i = \operatorname{rk} G_i$. The number of moduli arising from (a) is evidently

$$\sum_{i=0}^{p} (n_i^2 - 1)(g-1) + pg.$$

Indeed, the bundles G_i have degree $n_i\mu$ and the product of their determinants must be L. They are otherwise unconstrained. For (c), again using techniques in § 4, one sees that the number of moduli contributed by extensions is

$$\begin{aligned}\sum_{i=0}^{p} [h^1(G_i^* \otimes V_{i+1}) - 1] &\leq \sum_{i=0}^{p} [p - i - n_i(n_{i+1} + \cdots + n_p)(1-g)] - (p+1)\\ &= \frac{p(p+1)}{2} - \sum_{i=0}^{p-1} n_i(n_{i+1} + \cdots + n_p)(1-g) - (p+1)\\ &= \frac{(p+1)(p-2)}{2} - \sum_{i<j} n_i n_j (1-g).\end{aligned}$$

[2] One can be more rigorous. Identifying $Z^{\mathbb{P}} = \{x\} \times \mathbb{P}$ with $\mathbb{P}$ we see that restricting the universal exact sequence to $Z^{\mathbb{P}}$ gives us the quotient $\mathcal{O}_{\mathbb{P}} \otimes_{\mathbb{C}} V_x \twoheadrightarrow \mathcal{T}_0|Z^{\mathbb{P}}$. Let S be a scheme which has a quotient $\mathcal{O}_S \otimes_{\mathbb{C}} V_x \twoheadrightarrow \mathcal{L}$ on it where $\mathcal{L}$ is a line bundle. This quotient extends (uniquely) to a family parabolic structures $q_S^* V \twoheadrightarrow \mathcal{T}$ (on V) parameterized by S. By the Lemma, the kernel is a family of stable bundles. The universal property of the exact sequence (3.1) gives us a map $S \to \mathbb{P}$, and this map factors through $f^{-1}(V)$.

This gives

$$\begin{aligned}\operatorname{codim}(\mathbb{P} \setminus \mathbb{P}^s) &\geq \sum_{i<j} n_i n_j (g-1) - \frac{(p-1)(p+2)}{2}\\ &= B \qquad \text{(say)}.\end{aligned}$$

Now, $\sum_{i<j} n_i n_j \geq p(p+1)/2$, therefore

$$B \geq \frac{p(p+1)}{2}(g-1) - \frac{(p+2)(p-1)}{2}.$$

It follows that $B \geq 3$ whenever $p \geq 2$ and $g \geq 3$. If $p = 1$ and $n \geq 3$, then

$$B/(g-1) = \sum_{i<j} n_i n_j \geq 2$$

and one checks that $B \geq 3$ whenever $g \geq 3$.

Remark 5.1.1. We could use similar techniques to estimate $\operatorname{codim}(\mathcal{S} \setminus \mathcal{S}^s)$, but our task is made easier by the exact answers in [22, p. 48, A]. For just this remark, assume $d > n(2g-1)$, and let $a = (n, d)$. Then $a \geq 2$. Let $n_0 = n/a$. Then according to *loc. cit.*,

$$\operatorname{codim}(\mathcal{S} \setminus \mathcal{S}^s) = \begin{cases} (n^2-1)(g-1) - \dfrac{n^2}{2}(g-1) - 2 + g & \text{if } a \text{ is even} \\ (n^2-1)(g-1) - \dfrac{n^2+n_0^2}{2}(g-1) - 2 + g & \text{if } a \text{ is odd.} \end{cases}$$

It now follows that

$$\operatorname{codim}(\mathcal{S} \setminus \mathcal{S}^s) > 5$$

if n, g are in the range of Theorem 1.0.2(b).

6. Hodge theory

This section, which can be read independently of the rest of the paper, contains some results from Hodge theory that will be needed to complete the proofs of the main theorems.

6.1 *Purity*

We refer to [8] for the definition and basic properties of mixed Hodge structures. Deligne's fundamental result is that the cohomology groups of schemes with coefficients in $\mathbb{Z}$ carry canonical mixed Hodge structures. We will need certain purity results for these mixed Hodge structure for low degree cohomology of smooth open varieties. These results can be deduced by comparing ordinary cohomology to intersection cohomology and appealing to the work of Saito [20]. However we will give more elementary arguments, using a version of the Lefschetz hyperplane theorem.

The notation of this section will be independent of the others.

Lemma 6.1.1. *If Y is a smooth variety, Z a codimension k closed subscheme, and $U = Y \setminus Z$, then*

$$H^j(Y, \mathbb{Z}) \xrightarrow{\sim} H^j(U, \mathbb{Z})$$

for $j < 2k - 1$.

Proof. We have to show that $H^j_Z(Y, \mathbb{Z})$ vanishes for $j < 2k$. By Alexander duality (see for e.g. [12], p. 381, Theorem 4.7) we have

$$H^j_Z(Y, \mathbb{Z}) \xrightarrow{\sim} H_{2m-j}(Z, \mathbb{Z}),$$

where $m = \dim Y$ and H_* is Borel–Moore homology. Now use ([12], p. 406, 3.1) to conclude that the right side vanishes if $j < 2k$ (note that 'dim' in *loc. cit* is dimension as an analytic space, and in *op. cit.* it is dimension as a topological (real) manifold). □

Remark 6.1.1. In view of the above Lemma, it seems that Balaji's proof of Torelli (for Seshadri's desingularization of $\mathcal{SU}_X(2, \mathcal{O}_X)$) does not work for $g = 3$, for in this case, the codimension of $\mathbb{P} \setminus \mathbb{P}^s = 2$ (see [4], top of p. 624 and [3], Remark 9).

COROLLARY 6.1.1

$H^j(U, \mathbb{Z})$ is pure of weight j when $j < 2k - 1$.

We will need purity results even when compactification is not smooth. To this end we prove the following version of the Lefschetz theorem.

Theorem 6.1.1. *Let Y be an m-dimensional projective variety. Suppose that U is a smooth Zariski open subset. If H is a hyperplane section of Y such that $U \cap H$ is non-empty, then*

$$H^i(U, \mathbb{Q}) \to H^i(U \cap H, \mathbb{Q})$$

is an isomorphism for $i < m - 1$ and injective when $i = m - 1$.

Proof. We need some results involving Verdier duality. The standard references are [6] and [12]. Let S be an analytic space and p_S the map from S to a point. For $\mathcal{F} \in D^b_{\text{const}}(S, \mathbb{Q})$ (the derived category of bounded complexes of $\mathbb{Q}_S$-sheaves whose cohomology sheaves are $\mathbb{Q}_S$-constructible), set

$$D_S(\mathcal{F}) = \mathbb{R}\mathcal{H}\text{om}_S(\mathcal{F}, p^!_S\mathbb{Q}).$$

We then have by Verdier duality

$$\mathbb{H}^i(S, \mathcal{F}) \xrightarrow{\sim} \mathbb{H}^{-i}(S, D_S(\mathcal{F}))^*. \tag{6.1}$$

Here $\mathbb{H}^*$ denotes 'hypercohomology'.

For an open immersion $h: S' \hookrightarrow S$, one has canonical isomorphisms

$$\mathbb{R}h_* D_{S'}\mathcal{G} \xrightarrow{\sim} D_S(h_!\mathcal{G}) \tag{6.2}$$

and

$$\mathbb{R}h_! D_{S'}\mathcal{G} \xrightarrow{\sim} D_S(\mathbb{R}h_*\mathcal{G}). \tag{6.3}$$

Here $\mathcal{G} \in D^b_{\text{const}}(S', \mathbb{Q})$. The first isomorphism is easy (using Verdier duality for the map h) and the second follows from the first and from the fact that $D_{S'}$ is an involution. We have used (throughout) the fact that $h_!$ is an exact functor.

If S is smooth, we have

$$p^!_S\mathbb{Q} = \mathbb{Q}_S[2\dim S]. \tag{6.4}$$

In order to prove the theorem, let $V = U \setminus H$ and $W = Y \setminus H$. We then have a cartesian square

$$\begin{tikzcd} V \arrow[r, "\imath'"] \arrow[d, "\jmath'"'] & U \arrow[d, "\jmath,"] \\ W \arrow[r, "\imath"'] & Y \end{tikzcd}$$

where each arrow is the obvious open immersion. We have, by (6.2) and (6.3), the identity

$$\jmath_! \mathbb{R}\imath'_* D_V \mathbb{Q}_V = D_Y(\mathbb{R}\jmath_* \imath'_! \mathbb{Q}_V). \tag{6.5}$$

Consider the exact sequence of sheaves

$$0 \longrightarrow \imath'_! \mathbb{Q}_V \longrightarrow \mathbb{Q}_U \longrightarrow g_* \mathbb{Q}_{H \cap U} \longrightarrow 0,$$

where $g: H \cap U \rightarrow U$ is the natural closed immersion. It suffices to prove that $H^i(U, \imath'_! \mathbb{Q}_V) = 0$ for $i \leq m-1$. Now,

$$H^i(U, \imath'_! \mathbb{Q}_V) = \mathbb{H}^i(Y, \mathbb{R}\jmath_* \imath'_! \mathbb{Q}_V).$$

Using (6.1), (6.5) and (6.4), the above is dual to

$$\mathbb{H}^{-i}(Y, \jmath_! \mathbb{R}\imath'_* D_V \mathbb{Q}_V) = \mathbb{H}^{2m-i}(Y, \jmath_! \mathbb{R}\imath'_* \mathbb{Q}_V).$$

But $\jmath_! \mathbb{R}\imath'_* = \mathbb{R}\imath_* \jmath'_!$, and hence the above is

$$\begin{aligned} \mathbb{H}^{-i}(Y, \jmath_! \mathbb{R}\imath'_* D_V \mathbb{Q}_V) &= \mathbb{H}^{2m-i}(Y, \mathbb{R}\imath_*(\jmath'_! \mathbb{Q}_V)), \\ &= \mathbb{H}^{2m-i}(W, \jmath'_! \mathbb{Q}_V), \\ &= H^{2m-i}(W, \jmath'_! \mathbb{Q}_V). \end{aligned}$$

Now, W is an affine variety, and therefore, according to Artin [1], its constructible cohomological dimension is less than or equal to its dimension. Consequently, the above chain of equalities vanish whenever $i < m$ (see also [10]). □

We will use the notation of this theorem for the remainder of the section.

COROLLARY 6.1.2

Let $e = \operatorname{codim}(Y \setminus U)$. *For* $i < e-1$, *the Hodge structure* $H^i(U, \mathbb{Z})$ *is pure of weight* i.

Proof. This is true if U is projective. In general proceed using Bertini's theorem, induction, Theorem 6.1.1 and the fact that submixed Hodge structures of pure Hodge structures are pure [8]. □

6.2 *Polarizations*

Let Y, U, e, etc. be as in Theorem 6.1.1. Let $i \in \mathbb{N}$ and $\mathcal{L}$ a line bundle on Y be such that

(a) $H^j(U, \mathbb{Q}) = 0$ for $j = i-2, i-4, \ldots$ (note that this forces i to be *odd*);
(b) $i < e-1$;
(c) $\mathcal{L}$ is very ample.

Let M be the intersection of $k = m - e + 1$ hyperplanes in general position. Then M is a smooth variety contained in U. Let

$$l: H^i(U) \longrightarrow H_c^{2m-i}(U)$$

be the composite of

$$\begin{aligned} H^i(U) &\longrightarrow H^i(M) \\ &\longrightarrow H^{2m-2k-i}(M) \\ &\longrightarrow H_c^{2m-i}(U), \end{aligned}$$

where the first map is restriction, the second is 'cupping with $c_1(\mathcal{L})^{m-k-i}$' and the third is the Poincaré dual to restriction. The map l is also described as

$$x \longmapsto x \cup c_1(\mathcal{L})^{m-k-i} \cup [M].$$

One then has (easily).

Lemma 6.2.1. *If M' is another k-fold intersection of general hyperplanes, then $[M'] = [M]$. Therefore l depends only on $\mathcal{L}$.*

PROPOSITION 6.2.1

The pairing

$$\langle x, y \rangle = \int_U l(x) \cup y$$

on $H^i(U, \mathbb{C})$ gives a polarization on the pure Hodge structure $H^i(U, \mathbb{Z})$. This makes the associated complex torus $J^p(U)$ (where $i = 2p - 1$) into an abelian variety when $H^i(U)$ is of type $\{(p, p-1), (p-1, p)\}$.

Proof. By Theorem 6.1.1, we have an isomorphism

$$r: H^i(U) \longrightarrow H^i(M).$$

The latter Hodge structure carries a polarization given by

$$\langle \alpha, \beta \rangle = \int_M c_1(\mathcal{L})^{m-k-i} \cup \alpha \cup \beta.$$

The conditions on i and the Hodge–Riemann bilinear relations on the primitive part of $H^i(M, \mathbb{C})$, assure us that the above is indeed a polarization (see [11] or [25], Chap. V, § 6). Our conditions on i imply that the primitive part of $H^i(M)$ is everything so that the pth intermediate Jacobian of M is an abelian variety. The pairing on $H^i(M)$ translates to a polarization on $H^i(U)$ given by

$$\langle x, y \rangle = \int_U l(x) \cup y.$$

This gives the result. □

6.3 *Hodge structure of projective bundles*

Deligne's construction [8] gives a stronger result than is actually stated, namely that for a smooth quasi-projective variety X, $H^i(X)$ take values in the subcategory of polarizable mixed Hodge structures [5]. One pleasant feature of this subcategory is the following generalization of Poincaré reducibility.

Lemma 6.3.1. *The category of polarizable rational pure Hodge structures is semi-simple.*

Proof. This follows from [5]. □

COROLLARY 6.3.1

The category of polarizable rational pure Hodge structures satisfies cancellation, i.e. if $A \oplus B \cong A \oplus C$ *then* $B \cong C$.

PROPOSITION 6.3.1

Let $p : M \to N$ *be a* $(d-1)$*-fold fiber product of* $\mathbb{P}^{n-1}$*-bundles (which need not be locally trivial in the Zariski topology) over a smooth simply connected quasiprojective variety* N. *Then*

(a) $H^1(M,\mathbb{Z}) = 0$.
(b) $p^* : H^3(N,\mathbb{Z}) \to H^3(M,\mathbb{Z})$ *is surjective, and its kernel is finite. In particular* p^* *induces an isomorphism* $H^3(N,\mathbb{Z})_{\text{free}} \cong H^3(M,\mathbb{Z})_{\text{free}}$ *(where* $A_{\text{free}} = A/A_{\text{tors}}$*).*
(c) $H^i(M,\mathbb{Q}) = \oplus_{j\geq 0} H^{i-2j}(N,\mathbb{Q}) \otimes H^{2j}((\mathbb{P}^{n-1})^{d-1},\mathbb{Q})$ *as mixed Hodge structures.*

Proof. We will use the Leray spectral sequence with coefficients in $\mathbb{Z}$ and $\mathbb{Q}$. As N is simply connected, $R^ip_*\mathbb{Z}$ is the constant sheaf $H^i((\mathbb{P}^{n-1})^{d-1},\mathbb{Z})_N$. The first statement is an immediate consequence of the Leray spectral sequence since $H^1(N,p_*\mathbb{Z}) = 0$.

Note that aside from $H^3(N,p_*\mathbb{Z}) = H^3(N,\mathbb{Z})$, all the other E_2 terms that contribute to $H^3(M,\mathbb{Z})$ vanish. This implies that $H^3(M,\mathbb{Z})$ is a quotient of $H^3(N,\mathbb{Z})$. But it is an isomorphism after tensoring with $\mathbb{Q}$ by [7]. Since the quotient $H^3(N,\mathbb{Z}) \twoheadrightarrow H^3(M,\mathbb{Z})$ arising from the analysis of the spectral sequence is precisely p^*, therefore the kernel is finite. This implies that the induced map $H^3(N,\mathbb{Z})_{\text{free}} \to H^3(M,\mathbb{Z})_{\text{free}}$ is surjective with trivial kernel.

Suppose that $p : M \to N$ is a $\mathbb{P}^{n-1}$ bundle, that is $d = 2$. Let D be a 'nonvertical' divisor class on M i.e. a class which restricts nontrivially to each fiber of p (for instance $c_1(\omega_{M/N})$). Then the subspace generated by D^i can be identified with a copy of $\mathbb{Q}(-i)$ in $H^{2i}(M)$. Thus one gets a morphism of mixed Hodge structures,

$$\bigoplus_{j<n} H^{i-2j}(N)(-j) \to H^i(M)$$

given by summing the maps $\alpha \mapsto p^*\alpha \cup D^j$. Note that the restrictions of D^j generate the cohomology of each fiber. Therefore by the Leray–Hirch theorem the above maps are isomorphisms, and this proves the last statement in this case. In general, M is a fiber product of $\mathbb{P}^{n-1}$ bundles. Let D_k be the pullback of a nonvertical divisor class from the kth factor. Arguing as before, we get isomorphisms

$$\bigoplus_{j_1+\cdots j_{d-1}=j, j_i<n} H^{i-2j}(N)(-j) \to H^i(M)$$

obtained by summing the maps

$$\alpha \mapsto p^*\alpha \cup D_1^{j_1} \cup \ldots D_{d-1}^{j_{d-1}}.$$

□

By combining this with the previous corollary, and using the fact that sub-mixed Hodge structures of pure Hodge structures are pure, we obtain.

COROLLARY 6.3.2

Let $p_i : M_i \to N_i$ $i = 1,2$ *be two* $(\mathbb{P}^{n-1})^{d-1}$*-bundles satisfying the hypotheses on* p *in Proposition* 6.3.1. *If* $\phi : M_1 \to M_2$ *is a morphism of varieties inducing isomorphisms (of* *rational mixed Hodge structures)* $H^i(M_2, \mathbb{Q}) \to H^i(M_1, \mathbb{Q})$ *for* $i \leq k$, *and if these mixed Hodge structures are pure of weight i then for* $i \leq k$ *the Hodge structures* $H^i(N_2, \mathbb{Q})$ *and* $H^i(N_1, \mathbb{Q})$ *are pure of weight i and are noncanonically* isomorphic as rational pure Hodge structures of weight i.

7. Hodge structure on degree one moduli

Let $\mathcal{U}_X(n, 1)$ be the moduli space of semi-stable rank n vector bundles of degree 1 on X. There is a smooth morphism $\det : \mathcal{U}_X(n, 1) \to \mathrm{Pic}^1(X)$ given by the determinant, and the fiber over L' is just $\mathcal{SU}_X(n, L')$.

Atiyah and Bott ([2], 9.11) gave generators for the cohomology ring $H^*(\mathcal{U}_X(n, 1), \mathbb{Q})$ in terms of certain universal characteristic classes. As the restriction map on rational cohomology from $\mathcal{U}_X(n, 1)$ to $\mathcal{SU}_X(n, L')$ is surjective [*loc. cit.* 9.7], these give generators for the cohomology of the $\mathcal{SU}_X(n, L')$. We will describe these generators in a form which is convenient for us. Let E be a Poincaré vector bundle. Fix a line bundle $L' \in \mathrm{Pic}^1(X)$, let p_i denote the projections on $X \times \mathcal{SU}_X(n, L')$ and let c_r be the rth Chern class of E restricted to $X \times \mathcal{SU}(n, L')$. Let $\gamma_{r,i}$ denote the composition of the following morphisms of Hodge structures:

$$H^i(X, \mathbb{Q})(-r+1) \xrightarrow{p_1^*} H^i(X \times \mathcal{SU}_X(n, L'), \mathbb{Q})(-r+1),$$
$$H^i(X \times \mathcal{SU}_X(n, L'), \mathbb{Q})(-r+1) \xrightarrow{c_r \cup} H^{i+2r}(X \times \mathcal{SU}_X(n, L'), \mathbb{Q})(1),$$
$$H^{i+2r}(X \times \mathcal{SU}_X(n, L'), \mathbb{Q})(1) \xrightarrow{p_{2*}} H^{i+2r-2}(\mathcal{SU}_X(n, L'), \mathbb{Q}).$$

Theorem 7.0.1 [2]. *The ring* $H^*(\mathcal{SU}_X(n, L'), \mathbb{Q})$ *is generated by the images of the maps* $\gamma_{r,i}$ *for* $2 \leq r \leq n$ *and* $0 \leq i \leq 2$.

COROLLARY 7.0.3

For each i, any simple summand of the Hodge structure

$$H^i(\mathcal{SU}_X(n, L'), \mathbb{Q})$$

is, up to Tate twisting, a direct summand of a tensor power of $H^1(X)$. *The Hodge structure on* $H^i(\mathcal{SU}_X(n, L'), \mathbb{Q})$ *is independent of* $L' \in \mathrm{Pic}^1(X)$.

Proof. Let $P = \mathrm{Pic}^1(X)$. Let H be the direct sum of all the tensor products

$$H^{i_1}(X, \mathbb{Q})(-r_1+1) \otimes H^{i_2}(X, \mathbb{Q})(-r_2+1) \dots$$

indexed by the finite sequences $((i_1, r_1), (i_2, r_2), \dots)$ with

$$(i_1 + 2r_1 - 2) + (i_2 + 2r_2 - 2) + \cdots = i.$$

Then, by the theorem, there is a surjection $H \to H^i(\mathcal{SU}_X(n, L'), \mathbb{Q})$ given by the product of γ's. This implies the first statement. The above map extends to a surjection of variations of Hodge structures $H_P \to R^i\det_*\mathbb{Q}$, where the H_P denotes the constant variation with fiber H. The second statement now follows from the theorem of the fixed part ([8], 4.1.2). □

Since the relations among the above generators have recently been determined by Jeffrey and Kirwan [13], it is possible to make a complete determination of these Hodge structures (over $\mathbb{Q}$). It is not clear whether the maps $\gamma_{r,i}$ are surjective for integral cohomology, however one does have the following theorem.

Theorem 7.0.2 [18]. *The map*

$$\gamma_{2,1} : H^1(X,\mathbb{Z})(-1) \to H^3(\mathcal{SU}_X(n,L'),\mathbb{Z})$$

is an isomorphism.

This is not stated as such, but this is implicit in their proof of their third theorem.

8. Main theorems

8.1 *Natural polarizations*

We give a proof of the following 'folklore' theorem. The theta divisor and its multiples are the only natural polarizations on the Jacobian. To simplify the discussion, we work with polarizations in the Hodge theoretic sense. Let $\pi : \mathcal{X} \to T$ be a family of genus g curves over an irreducible base variety. Let θ be the standard polarization on $R^1\pi_*\mathbb{Z}$ corresponding to cup product, and let θ' be some other polarization.

Lemma 8.1.1. *If the canonical map to moduli space $T \to M_g$ is dominant, then there exist a positive integer m such that $\theta' = m\theta$.*

Proof. Let $t_0 \in T$ be a base point. θ_{t_0} can be viewed as a primitive vector in $V = \wedge^2 H^1 (X_{t_0},\mathbb{Z})^{\pi_1(T,t_0)}$. Therefore it is enough to prove that $V \otimes \mathbb{R}$ is one dimensional. After replacing T by a Zariski open subset, we can assume that the image $T' \subset M_g$ is disjoint from the locus of curves with automorphisms, and that $T \to T'$ is smooth. This guarantees (by Teichmuller theory) that $\pi_1(T')$ surjects onto the mapping class group which surjects onto $\mathrm{Sp}_{2g}(\mathbb{Z})$, and furthermore that the image of $\pi_1(T)$ has finite index in $\pi_1(T')$. Therefore $\pi_1(T)$ has Zariski dense image in $\mathrm{Sp}_{2g}(\mathbb{R})$. Consequently $V \otimes \mathbb{R} = \wedge^2 H^1(X_{t_0},\mathbb{R})^{\mathrm{Sp}_{2g}(\mathbb{R})}$ which is well-known to be spanned by the standard symplectic form. □

8.2 *First main theorem*

Theorem 8.2.1. *Let $\imath(n,g) = 2(n-1)g - (n-1)(n^2+3n+1) - 7$. Let X be a curve of genus $g \geq 2$. If $n \geq 4$ and $i < \imath(n,g)$ are integers, then for any pair of line bundles L, L' on X, the mixed Hodge structures $H^i(\mathcal{SU}^s_X(n,L),\mathbb{Q})$ and $H^i(\mathcal{SU}^s_X(n,L'),\mathbb{Q})$ are (non-canonically) isomorphic and are both pure of weight i.*

Proof. There is no loss of generality in assuming that $\deg L' = 1$. In fact, by 7.0.3, we can assume that L' is a specific line bundle, namely $L' = L(-D)$, where $D = x^1 + \cdots + x^{d-1}$. Then reverting to the our earlier notation, we have $\mathcal{S}_1 = \mathcal{SU}_X(n,L')$ and $\mathcal{S} = \mathcal{SU}_X(n,L)$. Consider the diagram

$$\mathcal{S}_1 \xleftarrow{\pi} \mathbb{P} \supset U \xrightarrow{f} U' \subset \mathcal{S}^s.$$

Then Corollaries 4.4.1, 4.4.2 and Lemma 6.1.1 implies

$$H^i(\mathbb{P},\mathbb{Q}) \cong H^i(U,\mathbb{Q}),$$
$$H^i(U',\mathbb{Q}) \cong H^i(\mathcal{S}^s,\mathbb{Q}),$$

for all $i < \imath(n,g)$. The theorem follows from Corollary 6.3.2 (see also Remark 3.3.1). □

8.3 *Second main theorem*

Theorem 8.3.1. *Let X be a curve of genus $g \geq 2$, $n \geq 2$ an integer and L a line bundle on X. Let $\mathcal{S}^s = \mathcal{SU}^s_X(n,L)$.*

(a) *If $g > (3/(n-1)) + ((n^2+3n+3)/2)$ and $n \geq 4$, then $H^3(\mathcal{S}^s,\mathbb{Z})$ is a pure Hodge structure of type $\{(1,2),\ (2,1)\}$, and it carries a natural polarization making the intermediate Jacobian*

$$J^2(\mathcal{S}^s) = \frac{H^3(\mathcal{S}^s,\mathbb{C})}{F^2 + H^3(\mathcal{S}^s,\mathbb{Z})}$$

into a principally polarized abelian variety. There is an isomorphism of principally polarized abelian varieties $J(X) \simeq J^2(\mathcal{S}^s)$.

(b) *If $\deg L$ is a multiple of n, then the conclusions of (a) are true for $g \geq 3$, $n \geq 2$ except the case $g = 3$, $n = 2$.*

Proof. We concentrate on part (a). The proof of part (b) is identical if we take the Hecke correspondence of § 5, and the codimension estimates of $\mathbb{P} \setminus \mathbb{P}^s$ and $\mathcal{S} \setminus \mathcal{S}^s$ given in that section (see 5.1). Consider the diagram

$$\mathcal{S}_1 \xleftarrow{\pi} \mathbb{P} \supset U \xrightarrow{f} U' \subset \mathcal{S}^s$$

once again. Then Corollaries 4.4.1, 4.4.2 and Lemma 6.1.1 imply

$$H^3(\mathbb{P},\mathbb{Z}) \cong H^3(U,\mathbb{Z}),$$
$$H^3(U',\mathbb{Z}) \cong H^3(\mathcal{S}^s,\mathbb{Z}).$$

$\mathcal{S}_1$ is unirational (see [22], pp. 52–53, VI.B), hence so is $\mathbb{P}$. Therefore these varieties are simply connected [21]. Since $\operatorname{codim}(\mathbb{P} \setminus U) > 1$ it follows that U is simply connected (purity of branch locus). The homotopy exact sequence tells us that U' is simply connected. Proposition 6.3.1 applied to π and f implies that on the third cohomology, π^* and f^* are surjective with finite kernels. Since π is locally trivial in the Zariski topology, π^* is in fact an isomorphism. Combining these facts with the isomorphisms above produces a map of Hodge structures

$$H^3(\mathcal{S}^s,\mathbb{Z}) \to H^3(\mathcal{S}_1,\mathbb{Z})$$

which is surjective with finite kernel. As an immediate consequence we have

$$H^3(\mathcal{S}_1,\mathbb{Z})_{\text{free}} \cong H^3(\mathcal{S}^s,\mathbb{Z})_{\text{free}}.$$

This, together with Theorem 7.0.2, yields an isomorphism of Hodge structures:

$$H^1(X,\mathbb{Z})(-1) \cong H^3(\mathcal{S}^s,\mathbb{Z})_{\text{free}} \tag{8.1}$$

which yields an isomorphism of tori $J(X) \cong J^2(\mathcal{S}^s)$.

The next step is to construct a polarization on $H^3(\mathcal{S}^s,\mathbb{Z})$. One knows from the results of Drezet and Narasimhan [9], that $\operatorname{Pic}(\mathcal{S}^s) = \mathbb{Z}$ (see p. 89, 7.12 (especially the proof) of *loc.cit.*). Moreover, $\operatorname{Pic}(\mathcal{S}) \to \operatorname{Pic}(\mathcal{S}^s)$ is an isomorphism. Let ξ' be the ample generator of $\operatorname{Pic}(\mathcal{S}^s)$. It is easy to see that there exists a positive integer r, independent of (X,L) (with genus $X = g$), such that $\xi = \xi'^r$ is very ample on $\mathcal{S}$ (we are not distinguishing between line bundles on $\mathcal{S}^s$ and their (unique) extensions to $\mathcal{S}$). Embed $\mathcal{S}$ in a suitable projective space via ξ. Let $e = \operatorname{codim}(\mathcal{S} \setminus \mathcal{S}^s)$. Let M be the intersection of $k = \dim \mathcal{S} - e + 1$ hyperplanes (in general position) with $\mathcal{S}^s$. Then M is smooth, projective and contained in $\mathcal{S}^s$. Let $p = \dim \mathcal{S}$ and H_c^* — cohomology with compact support. We then have a map

$$l\colon H^3(\mathcal{S}^s) \longrightarrow H_c^{2p-3}(\mathcal{S}^s)$$

defined by

$$x \longmapsto x \cup c_1(\xi)^{p-k-3} \cup [M].$$

If M' is another k-fold intersection of general hyperplanes, then $[M'] = [M]$. Hence l depends only on ξ. According to Proposition 6.2.1, the pairing on $H^3(\mathcal{S}^s, \mathbb{C})$ given by

$$\langle x, y\rangle = \int_{\mathcal{S}^s} l(x) \cup y$$

gives a polarization on the Hodge structure of $H^3(\mathcal{S}^s)$. Pulling this back via the isomorphism (8.1) gives a second polarization on $H^1(X)$. If we can show that $\langle\, , \rangle$ varies over the whole M_g, then we can appeal to Lemma 8.1.1 to show that there exists a positive integer m (independent of X) such that $1/m\langle\, , \rangle$ coincides with the standard principal polarization of $H^1(X)$, and this will complete the proof.

Let T be the moduli space parameterizing tuples

$$(X, x^1, \ldots, x^{d-1}, L, \lambda)$$

consisting of a genus g curve, $d-1$ distinct points, a degree d line bundle and a level 3 structure. T is an irreducible variety which surjects onto M_g. The inclusion of the level structure guarantees that T is fine, and therefore there is a universal curve $\mathcal{X} \to T$ together with $d-1$ sections and so on. All of the constructions given so far can be carried out relative to T, in other words we can construct a diagram of T-schemes

$$\begin{tikzcd} S_1 \arrow[dr, "p_1"'] & P \arrow[l] \arrow[r, dashed] \arrow[d] & S \arrow[dl, "p"] \\ & T & \end{tikzcd}$$

whose fibers are the Hecke correspondences. Furthermore there is polarization on $R^3p_*\mathbb{Z} \cong R^1p_{1*}\mathbb{Z}(-1)$, which restricts to the one constructed in the previous paragraph. We are now in a position to apply Lemma 8.1.1 to conclude the proof. □

Acknowledgements

We wish to thank Profs M S Narasimhan, C S Seshadri and S Ramanan for their encouragement and their help. Thanks to V Balaji, L Lempert, N Raghavendra and P A Vishwanath for helpful discussions. Balaji made us aware of the problem, and generously discussed his proof (in [4]) of the Torelli theorem for Seshadri's desingularization of $\mathcal{SU}_X(2, \mathcal{O}_X)$. We wish to thank the anonymous referee for the care with which he/she read the paper, and for the helpful comments. The second author gratefully acknowledges the four wonderful years he spent at the SPIC Science Foundation, Chennai. (DA) was partially supported by the NSF.

References

[1] Artin M, *Théorème de finitude pur un morphisme propre: dimension cohomologique des schemes algébriques affines*, SGA-4, expose XIV, *Lecture Notes in Math.* (New York: Springer-Verlag, Berlin-Heidelberg) (1973) vol. 305

[2] Atiyah M and Bott R, The Yang-Mills equations over Riemann surfaces, *Phil. Trans. R. Soc. London* **A308** (1982)

[3] Balaji V, Cohomology of certain moduli spaces of vector bundles. *Proc. Indian Acad. Sci. (Math. Sci.)* **98** (1987) 1–24

[4] Balaji V, Intermediate Jacobian of some moduli spaces of vector bundles on curves. *Amer. Jour. of Math.* **112** (1990) 611–630

[5] Beilinson A, Notes on absolute Hodge cohomology. Applications of algebraic K-theory tc algebraic geometry and number theory, *Amer. Math. Soc.* (1986)

[6] Borel A *et al*, Intersection Cohomology, *Progress in Mathematics* (Basel: Birkhauser) (1984) vol. 50

[7] Deligne P, Théorème de Lefschetz et critères de dégénérescence de suites spectrales, *Publ Math. I.H.E.S* **35** (1968) 107–126

[8] Deligne P, Théorie de Hodge II. *Publ. Math. I.H.E.S* **40** (1972) 5–57

[9] Drezet J M and Narasimhan M S, Groupe de Picard des variétés de modules de fibré semistables sur les courbes algébriques, *Invent. Math.* **97** (1989) 53–95

[10] Goresky M and MacPherson R, Intersection Homology II. *Invent. Math.* **72** (1983) 77–129

[11] Griffiths P A, Periods of integrals on algebraic manifolds, I, II, *Amer. J. Math.* **90** (1968) 568-626, 805–865

[12] Iversen B, *Cohomology of Sheaves* (Universitext. Springer-Verlag, Berlin-Heidelberg-Nev York-Tokyo) (1986)

[13] Jeffrey L and Kirwan F, Intersection theory on moduli spaces of holomorphic bundles o arbitrary rank on a Riemann surface, preprint (1996)

[14] Kouvidakis A and Pantev T, The automorphism group of the moduli space of semi stabl vector bundles, *Math. Ann.* **302** (1995) 225–268

[15] Mehta V B and Seshadri C S, Moduli space of parabolic vector bundles on curves, *Math Annalen* **248** (1980) 205–239

[16] Mumford D and Newstead P E, Periods of a moduli space of vector bundles on curves, *Ame J. Math.* **90** (1968) 1201–1208

[17] Narasimhan M S and Ramanan S, Geometry of Hecke Cycles-I. In *C. P. Ramanujam— Tribute* (Oxford University Press) (1978) TIFR Bombay

[18] Narasimhan M S and Ramanan S, Deformations of a moduli space of vector bundles, *Ann. c Math.* **101** (1975) 391–417

[19] Ramanan S, The moduli space of vector bundles over an algebraic curve, *Math. Annalen* **20** (1973) 69–84

[20] Saito M, Modules de Hodge polarizables, *RIMS* (1988)

[21] Serre J P, On the fundamental group of a unirational variety, *J. London Math. Soc.* **34** (1968 481–484

[22] Seshadri C S, Fibrés vectoriels sur les courbes algébriques, *Asterisque.* (Société mathématiqu de France) (1982) vol. 96

[23] Seshadri C S, Desingularisations of moduli varieties of vector bundles on curves in *Int. Symp on Algebraic Geometry* (ed.) M Nagata (1977) (Kyoto) pp. 155–184

[24] Tyurin A N, The geometry of moduli of vector bundles, *Russ. Math. Surveys* **29** (1974) 57–8

[25] Wells R O, *Differential Analyisis on Complex manifolds* (New York: Springer-Verlag) (198(

Proc. Indian Acad. Sci. (Math. Sci.), Vol. 110, No. 1, February 2000, pp. 27–34.
© Printed in India

An intrinsic approach to Lichnerowicz conjecture

AKHIL RANJAN

Department of Mathematics, Indian Institute of Technology, Mumbai 400 076, India
Email: aranjan@math.iitb.ernet.in

MS received 19 February 1999

Abstract. In this paper we give a proof of Lichnerowicz conjecture for compact simply connected manifolds which is intrinsic in the sense that it avoids the *nice embeddings* into eigenspaces of the Laplacian. Even if one wants to use these embeddings, this paper gives a more streamlined proof. As a byproduct, we get a simple criterion for a polynomial to be a Jacobi polynomial.

Keywords. Harmonic manifolds; Blaschke manifold; mean curvature; Jacobi differential equation; Ricci curvature; compact rank one symmetric spaces; nice embeddings.

1. Introduction

The object of this paper is to present an intrinsic proof of the Lichnerowicz's conjecture for the compact simply connected harmonic manifolds. For the definition of harmonic manifolds, see [1]. One of the characterizations is that the geodesic spheres around any point have constant mean curvature depending only on the radius of the sphere. Alternatively, there exist eigenfunctions of the Laplacian which depend only on the distance from the point, the so called radial eigenfunctions. It suffices to consider only small values of the radii. Lichnerowicz showed that for dimension less than or equal to 4, such a manifold must be either flat or a locally symmetric space of rank one (see [5, 1]). He quite naturally asked whether the same was true in higher dimensions. Great progress was made in the case of compact simply connected harmonic manifolds, a detailed account of which is given in Besse's book [1]. It is shown that these are all Blaschke manifolds and their Ricci tensor is proportional to the metric tensor or in other words they are Einstien manifolds. From topological point of view each such manifold has its (integral) cohomology ring isomorphic to precisely one of the compact rank one symmetric space to be referred to as its *model CROSS* henceforth. This result is due to Allemigeon. A particularly striking discovery about compact harmonic spaces is a family of isometric minimal immersions into the round spheres in eigenspaces of the Laplacian acting on the space of square-integrable functions. Moreover, any two geodesics were shown to be congruent to each other under some Euclidean isometry. These are now known as *Besse's nice embeddings* or helical immersions [12]. In 1990, Szabo [11] successfully used them along with other known facts about harmonic manifolds to answer Lichnerowicz's query affirmatively for compact simply connected harmonic manifolds. In contrast to compact case, Damek and Ricci in 1992 [4] produced a family of examples of homogeneous harmonic manifolds which are not locally symmetric. In Szabo's paper the key point was to show that the volume function of a compact simply connected harmonic manifold when expressed in terms of normal coordinates coincided with that of its *model CROSS*. To this end he goes through the following steps:

1. He establishes what he calls the *basic commutativity in harmonic spaces*. This implies in conjunction with Allamigeon's theorem that for any point p on the manifold and any eigenvalue λ of the Laplacian there exist eigenfunctions which depend only on radial distance from p. Moreover, starting with any eigenfunction and averageing over geodesic spheres around p we get such a radial eigenfunction.
2. By moving the point p along a geodesic and averaging in the said manner we get a parallely displaced family of functions in the eigenspace which is finite dimensional. This along with the fact that each geodesic is periodic of period *assumed* to be 2π enables one to conclude that the *radial* eigenfunctions alluded to above are polynomials in cosine of the radial distance.
3. At this stage the *nice embeddings* are used to pin down the volume function in geodesic normal coordinates.
4. Finally it follows that the first radial eigenfunction is linear of the form $A\cos r + B$ and studying the nice embedding in the first eigenspace one shows symmetry easily.

In this paper we show that the *nice embeddings* can be avoided in the step (3) above. Steps (1) and (2) do not require them anyway. Our proof of step (3) can be regarded as a streamlined version of that given by Szabo. As for the last step one can either use *nice embeddings* or the partial solution to a problem of Antonio Ros about the first eigenvalue of P-manifolds given in [8, 9]. Alternatively, one can use induction to complete the proof. In both the latter cases one has worked wholly within the manifold thereby proving Szabo's theorem intrinsically. It is also worth noting that the proof is over as soon as the step (3) is completed in those cases where the manifold is spherical by the Lichnerowicz–Obata's theorem or the Bonnet–Meyers' theorem.

2. Laplacian on radial functions

Let $\Theta(r)$ denote the volume function on a simply connected compact harmonic manifolc M in terms of normal coordinates and $\sigma(r)$ the mean curvature of any geodesic sphere o radius r. It is easily shown that

$$\frac{\Theta'(r)}{\Theta(r)} = \sigma(r).$$

Further for a point p and an eigenvalue λ of the Laplacian Δ, let u be an eigenfunctioı which depends only on radial distance r from p. As shown by Szabo ([11], p. 8, eq. (2.1) u satisfies the following ODE

$$u'' + \sigma(r)u' + \lambda u = 0. \qquad (2.1$$

Here $'$ means derivative with respect to r. We would like to study how closely Θ and agree with their anologues in its *model CROSS*. Let us first define the volume function o all of *real* line as follows. Consider a geodesic γ through a point p. Let $J_2, \ldots, J_d$ be th Jacobi fields along γ which vanish at $\gamma(0)$ and whose derivatives at $\gamma(0)$ form a orthonormal basis along with $\gamma'(0)$. Let $E_1, \ldots, E_d$ be parallel translation of the abov orthonormal basis along γ, E_1 being $\gamma'(r)$. Now set

$$\Theta(r) = \langle J_2 \wedge \cdots \wedge J_d,\ E_2 \wedge \cdots \wedge E_d \rangle(r).$$

By virtue of it being a Blaschke manifold, Θ when considered as a function on whole of the real line has the following properties:

1. It is periodic of period 2π.
2. It has zeroes of order $k-1$ at $r = n\pi$ for n any odd integer and zeroes of order $d-1$ at $r = n\pi$ for n any even integer. Here d is the dimension of M and k is the degree of the generator of the cohomology ring of M.
3. $\Theta(r) = (-1)^{d-1}\Theta(-r)$.

This clearly allows us to write

$$\Theta(r) = \mathrm{e}^{\alpha(\cos r)} \sin^{d-1}(r/2) \cos^{k-1}(r/2)$$

or $\Theta(r) = \mathrm{e}^{\alpha(\cos r)}\Theta_0(r)$ where $\Theta_0(r)$ is the volume function of the *model CROSS* and α is a smooth (actually analytic) function on $[-1, 1]$ with $\alpha(1) = 0$. Hence $\sigma(r) = \sigma_0(r) - \sin(r)\alpha'(\cos r)$ since $\sigma = \Theta'/\Theta$.

Caution: The conventional volume function is the absolute value of the one we have defined. They both agree within the *injectivity radius* i.e. for $0 \leq r \leq \pi$. For $0 < r < \pi$, an easy calculation gives that

$$\sigma_0(r) = \frac{1}{2\sin r}[(d-1)(1+\cos r) - (k-1)(1-\cos r)].$$

By Lemma 4.2 of [11], u in eq. (2.1) is of the form $u = f(\cos r)$ for some *polynomial* f. Inserting all this data in 2.1 and setting $\cos r = x$ we see that f satisfies the following:

$$(1-x^2)f'' - \left[\frac{d}{2}(1+x) - \frac{k}{2}(1-x) + (1-x^2)\alpha'(x)\right] f' + \lambda f = 0, \quad -1 \leq x \leq 1. \tag{2.2}$$

In the above equation $'$ denotes derivative with respect to x.

3. Jacobi differential equation

The differential equation

$$(1-x^2)u'' - [(1+b)(1+x) - (1+a)(1-x)]u' + \lambda u = 0 \tag{3.3}$$

has been studied classically as a (singular) Sturm–Liouville equation on $[-1, 1]$ and it is known that for a and b in $(-1, \infty)$ and under natural boundary conditions (u bounded as $|x| \to 1$) solutions exist for $\lambda = n(n+a+b+1), n \in \mathbb{N}$ and for each such value of λ, u is a polynomial of degree n. Moreover, u is unique upto a scalar multiple. In fact these polynomials form a complete orthogonal system in $L^2([-1, 1], \rho dx)$ where $\rho(x) = (1+x)^a(1-x)^b$ is the weight function. These are known as Jacobi polynomials (see [2], p. 289). This differential equation is known as Jacobi differential equation with parameters a and b. We assume that $a, b > -1$.

In this section we consider a *perturbed* Jacobi equation where we have an extra term of the form $(1-x^2)\delta(x)$ as a coefficient of u', δ being a continuous function on $[-1, 1]$. Comparing with the corresponding equation satisfied by the polynomial f in the previous section we easily see that $1+b = d/2$, $1+a = k/2$ and $\delta = \alpha'$. We also know that k cannot exceed d and can only take values in $2, 4, 8, d$, hence $a, b > -1$ is clearly true. Now we are ready to state our main theorem.

Theorem 3.1. *If the perturbed Jacobi differential equation*

$$(1-x^2)u'' - [(1+b)(1+x) - (1+a)(1-x) + (1-x^2)\delta(x)]u' + \lambda u = 0 \tag{3.4}$$

admits a nonconstant polynomial as a solution for some value of λ, then the perturbatio term δ must vanish identically.

COROLLARY 3.1

The perturbation term α' in 2.2 vanishes. Consequently α is identically zero and henc $\sigma = \sigma_0$ as well as $\Theta = \Theta_0$.

The proof of the above will be broken into two lemmas. Let P be a polynomial whic we assume to be nonconstant and monic which satisfies 3.4 for a suitable λ.

Lemma 3.1. *δ must be a rational function with the degree of the numerator being strictl less than that of the denominator.*

Proof.

$$\delta = \frac{LP + \lambda P}{(1-x^2)P'},$$

where

$$L = (1-x^2)\frac{\mathrm{d}^2}{\mathrm{d}x^2} - [(1+b)(1+x) - (1+a)(1-x)]\frac{\mathrm{d}}{\mathrm{d}x}$$

is the Jacobi differential operator (with parameters a and b). Clearly both numerator an denominator of δ are polynomials with denominator nonvanishing and of degree strictl more than that of the numerator. Hence the claim. ■

Lemma 3.2. *Let $\delta = p/q$ as a quotient of relatively prime polynomials with q bein monic. Then*

1. *All the roots of q are simple and in $\mathbb{C} \setminus [-1,1]$.*
2. *$q|P'$ and $q|P$.*
3. *Let $q = \prod(x-\beta_i)$ and m_i be the multiplicity of $\beta_i \in P$, then $m_i \geq 2$.*
4. *If we put $q_1 = \prod(x-\beta_i)^{m_i - 1}$, then $\delta = q_1'/q_1$.*
5. *± 1 cannot be roots of P.*
6. *Roots of P which are not common with those of q, are all simple.*

Proof. Let $P = \prod(x-\beta_i)^{m_i}$ where β_i are distinct complex numbers and m_i are natur numbers which are nonzero. Let

$$v = \frac{P'}{P} = \sum \frac{m_i}{x-\beta_i}. \tag{3.5}$$

Then v satisfies the Riccati equation (see [2], p. 124)

$$v' + v^2 = \left[\frac{1+b}{1-x} - \frac{1+a}{1+x} + \delta(x)\right]v - \frac{\lambda}{1-x^2}. \tag{3.6}$$

From 3.5 we get

$$v' + v^2 = \sum_i \frac{m_i^2 - m_i}{(x-\beta_i)^2} + \sum_{i \neq j} \frac{2m_i m_j}{(\beta_i - \beta_j)(x-\beta_i)}. \tag{3.7}$$

In the above equation we have expanded the *cross terms* occuring in v^2 into partial fractions. Now let $q = \prod (x-\alpha_j)^{r_j}$ where α_j are distinct complex numbers and $r_j \geq 1$. Since p and q are relatively prime, if we expand $\delta = p/q$ in partial fractions, $1/(x-\alpha_j)^{r_j}$ will survive for each j. They will continue to survive after multiplication by $v = \sum m_i/(x-\beta_i)$ and further expansion into partial fractions. Now if we compare the rhs of (3.6) and (3.7) after expanding into partial fractions we find that we must have $\{\alpha_j\} \subset \{\beta_i\}$ and $r_j = 1$ for each j. This proves that the roots of q are simple. Since $\delta = p/q$ is continuous on $[-1, 1]$ roots of q must be away from $[-1, 1]$. This poves the first assertion.

From the second assertion $q|P$ is clear from above. To show that $q|P'$, we note that $LP + \lambda P = (p(1-x^2)P')/q$ is a polynomial. Hence $q|P'$ since it is relatively prime to p, $1-x$, and $1+x$. This gives the second claim.

Put $S = \{\beta_i\}, S' = \{\alpha_j\}$, then $S' \subset S$. Also put $S'' = S \setminus S'$. We can then write $q = \prod_{S'}(x-\beta_j)$ and hence $\delta = \sum_{S'} c_i/(x-\beta_i)$ where c_i are nonzero numbers. Coming back to the third statement, let us compare the coefficient of $1/(x-\beta_i)^2$ in the rhs of (3.6) and (3.7) (after partial fractions) for $\beta_i \neq \pm 1$. We see that

$$c_i\, m_i = m_i^2 - m_i \text{ for } i \text{ s.t } \beta_i \in S' \text{ and } m_i^2 - m_i = 0 \text{ if } \beta_i \in S'' \setminus \{\pm 1\}.$$

From this we can conclude that

$$c_i = m_i - 1 \text{ for } i \text{ s.t. } \beta_i \in S'$$

and $m_i = 1$ for i s.t. $\beta_i \in S'' \setminus \{\pm 1\}$ (since $m_i \neq 0$ for each i).

$$\frac{p}{q} = \sum_{S'} \frac{m_i - 1}{x - \beta_i}.$$

The first of these shows that $m_i \geq 2$ for $\beta_i \in S'$ because $c_i \neq 0$ and the second one can be rewritten as $p/q = q_1'/q_1$, where $q_1 = \prod_{S'}(x-\beta_i)^{m_i - 1}$. These are just the third and the fourth assertions.

For the fifth claim, write

$$P(x) = \prod_S (x-\beta_i)^{m_i} = (x-1)^A (x+1)^B \prod_{S'} (x-\beta_i)^{m_i} \prod_{S'' \setminus \{\pm 1\}} (x-\beta_i).$$

Comparing coefficients of $(x-1)^{-2}$ and $(x+1)^{-2}$ we see that

$$A^2 - A = -(1+b)A \text{ and } B^2 - B = -(1+a)B.$$

Since a and b are more than -1 and A and B are natural numbers we get $A = B = 0$. Thus $S = S' \cup S''$; the set of roots of P, is disjoint from $\{\pm 1\}$.

Finally, $S'' \setminus \{\pm 1\} = S''$ so that for $\beta_i \in S''$ we have $m_i = 1$ which is the sixth assertion. ■

4. Proof of theorem 3.1

We have the following facts: $P(x) = \prod_{S'}(x-\beta_i)^{m_i} \prod_{S''}(x-\beta_i)$, $m_i \geq 2 . q_1(x) = \prod_{S'}(x-\beta_i)^{m_i-1}$. Clearly, $q_1|P'$. Let $P'/q_1 = n\prod_j (x-\gamma_j)^{M_j} = R(x)$, say. (Here $n = \deg P$.)

Then γ_j are those roots are P' which are not common with those of P. This is so becaus the roots of P in S'' are all simple. Hence

$$\{\gamma_j\}\bigcap S = \emptyset = S\bigcap\{\pm 1\}. \qquad (4.8$$

Substituting the expressions obtained for P, P' and δ in the equation (3.4), dividing b $(1-x^2)P'$ and simplifying we get

$$\sum \frac{M_j}{x-\gamma_j} + \frac{a+1}{x+1} + \frac{b+1}{x+1} = -\frac{\lambda \prod_S (x-\beta_i)}{n \prod (x-\gamma_j)^{M_j}}. \qquad (4.9$$

Putting $x = \beta \in S$ and using (4.8) we find that

$$\sum_j \frac{M_j}{\beta-\gamma_j} + \frac{a+1}{\beta+1} + \frac{b+1}{\beta-1} = 0, \quad \text{for each such } \beta. \qquad (4.10$$

This in turn implies that $\beta \in \text{conv}\{\gamma_j, \pm 1\}$ (where conv denotes the *convex hull*). O $S \subset \text{conv}\{\gamma_j, \pm 1\}$. On the other hand by Lucas' theorem $\text{conv}\{\gamma_j\} \subset \text{conv}\, S$. Now th argument of Lemma 4.6 ([11], p. 23) goes through and shows that $S \subset (-1, 1)$. Also a observed earlier S' is disjoint from $[-1, 1]$. Therefore, S' must be empty so that q i constant 1 and hence $p = \delta$ vanishes identically as $\deg(p) < \deg(q)$. ∎

An interesting corollary is the following charcterization of the Jacobi polynomial which should be of theoretical interest.

COROLLARY 4.2

If P is a nonconstant polynomial such that for some complex number λ the rationa function $(P'' + \lambda P)/(1-x^2)P'$ has simple poles at ± 1 with positive residues $1+a, 1+$ respectively and no other pole in the real interval $(-1, +1)$, then P is a Jacobi pol nomial with parameters a and b. ∎

5. Proof of the conjecture

Let us first recall a theorem of Antonio Ros ([9], Theorem 4.2, p. 402).

Theorem 5.2. *Let M be an n-dimensional $P_{2\pi}$-manifold and suppose that the Ric tensor, S, and the metric, $\langle , \rangle$, on M verify the relation $S \geq k\langle , \rangle$, where k is a real constan Let λ_1 be the first eigenvalue of the Laplacian of M. Then we have*

$$\lambda_1 \geq \frac{1}{3}(2k + n + 2).$$

The proof is short and elegant and independent of the other computations done in h paper. It also follows that, in case the equality holds, any first eigenfunction f has th property that for any geodesic γ_u, $f(\gamma_u(t)) = A_u \cos t + B_u \sin t + C_u$. This property wi be referred to as Ros' property in what follows. Also here and in the sequel γ_u will alwa denote the geodesic with initial condition u.

Ros also noted that for *CROSSES* the equality holds. He naturally asked as to wh restrictions apply to M if equality held.

As a partial answer to the above question the following theorem has been proved in [and [9]:

Theorem 5.3. *If equality holds in the Ros' estimate for λ_1 of a P-manifold and if M admits a corresponding eigenfunction without saddle points, then M is a CROSS.*

Also see [7] for another related result.

Proof of the conjecture. Now from corollary 3.1 to our main theorem $\Theta = \Theta_0$ and hence $\mathrm{Ric}_M = \mathrm{Ric}_0$ and $\lambda_1(M) = \lambda_1(M_0)$ where M_0 denotes the *model CROSS*. Moreover, from any point on M the first radial eigenfunction is of the form $\cos r + C$ and hence without saddle points. The proof of the Lichnerowicz conjecture is now complete. ■

It is however, possible to avoid the use of this theorem and give a more elementary inductive proof of the conjecture. It should be noted that in the spherical case the proof is already over by the well known Lichnerowicz–Obata theorem [6] or more easily by the rigidity part of the Bonnet–Meyers' theorem [3]. So let M be of projective type.

Claim I. For any $p \in M$, $C(p)$- the cut locus of p is totally geodesic.

Proof. Let $f = \cos r + C$ be a radial first eigenfunction from p, where $r = d(p, .),$. Let $u \in T_qC(p)$ be a unit vector. Then $\langle \nabla f, u\rangle = 0$ and $\langle (\nabla^2 f)(u), u\rangle = 0$, where ∇^2 denotes the Hessian. Hence by Ros' property $f(\gamma_u(t)) = f(\gamma_u(0)) = f(C(p))$ for every t. It follows that γ_u lies in the level set $C(p)$. ■

Claim II. $C(p)$ is itself harmonic.

Proof. Let $q \in C(p)$ and h be a radial first eigenfunction from q. Further, let σ be a geodesic in $C(p)$ starting at q and $q' = \sigma(s_0)$ be a point on it. Let $N_{q'}$ be the normal space to $T_{q'}C(p)$. For any unit vector $v \in N_{q'}$, $\gamma_v(\pi) = p$, and hence $h(\gamma_v(\pi))$ is independent of v. Now ∇h at q' points in the direction of $\sigma'(s_0)$ and thus is perpendicular to v. It follows that $h(\gamma_v(t)) = A_v \cos t + C_v$, since $B_v = 0$. Moreover,

$$h(q') = h(\gamma_v(0)) = A_v + C_v = \cos s_0 + C$$
$$h(p) = h(\gamma_v(\pi)) = -A_v + C_v = -1 + C, \quad \text{as } d(p,q) = \pi$$

Hence, it follows that A_v depends only on $s_0 = d(q, q')$ and is equal to the value it takes in the model CROSS in a parallel situation. (Observe that the constant C is same as in the model.) Let Δ_c denote the Laplacian in the cut locus $C(p)$. At the point q',

$$\Delta_c h = \Delta h - \mathrm{tr}_{N_{q'}}(\nabla^2 h).$$

This implies that

$$\Delta_c h = \lambda_1(M_0)h - \sum A_{v_i},$$

where $\{v_i\}$ form an orthonormal basis of $N_{q'}$. If $\cos r + C'$ gives the radial first eigenfunction in the cut locus of the model CROSS M_0 with eigenvalue $\lambda_{1,c}$, it follows that on setting $h' = h + (C' - C)$, it becomes a radial eigenfunction in $C(p)$ of eigenvalue $\lambda_{1,c}$. Thus $C(p)$ is a harmonic manifold as claimed. ■

Claim III. M is a symmetric space.

Proof. By induction hypothesis, $C(p)$ is a CROSS of sectional curvatures lying in $[\frac{1}{4}, 1]$. Here we have assumed that M is not spherical so that $C(p)$ is not a point. Now let c be any

geodesic in $C(p)$ and $\mathcal{R}$ be the curvature operator along c given by $\mathcal{R}$
Obviously, it leaves $T_{c(t)}C(p)$ and $N_{c(t)}$ invariant. These bundles are also par
This forces the Jacobi fields along c which vanish at $c(0)$ and whose initial de
$N_{c(0)}$ to remain in N throughout. Since the trace of $\mathcal{R}$ gives the Ricci curvatu
that tr $\mathcal{R}|N$ equals $k/4$ where k is the degree of the generator of $H^*(M,Z)$.
rank of the normal bundle N of $C(p)$. Applying the Bonnet–Meyers' met
normal Jacobi fields and noting that these cannot vanish again before 2π
complement of $(k-1)$ fields which vanish at π is tangent to $C(p)$) we find
these must be of the form $\sin(t/2)E$ for some parallel field E. Hence all the
along c are as in the model CROSS. Finally, since any geodesic can be m
some cut locus, symmetry of M follows.

References

[1] Besse A L, *Manifolds all of whose geodesics are closed* (Berlin: Springer) (1
[2] Birkhoff G and Rota G C, *Ordinary differential equations* (John Wiley and S
[3] Cheeger J and Ebin D, *Comparison Theorems in Riemannian Geometry* (N
(1975)
[4] Damek E and Ricci F, A class of nonsymmetric harmonic Riemmanian spaces
(1992) 139–142
[5] Lichnerowicz A, Sur les espaces riemanniens complétement harmoniques, *B*
France **72** (1944) 146–168
[6] Obata M, Certain conditions for a Riemannian manifold to be isometric to a s
Soc. Jpn. **14** (1962) 330–340
[7] Ranjan A and Santhanam G, A generalisation of Obata's theorem, *J. Geom.*
357–375
[8] Ranjan A and Santhanam G, The first eigenvalue of P-manifolds, *Osaka J. M*
821–842
[9] Ranjan A and Santhanam G, Correction to "The first eigenvalue of P-manifold
Osaka J. Math.
[10] Ros A, Eigenvalue inequalities for minimal submanifolds and P-manifolds,
(1984) 393–404
[11] Sakamoto K, Helical immersions into a unit sphere, *Math. Ann.* **261** (1982) 6
[12] Szabo Z I, The Lichnerowicz conjecture on harmonic manifolds, *J. Diff. Geom.*
28

Proc. Indian Acad. Sci. (Math. Sci.), Vol. 110, No. 1, February 2000, pp. 35–53.
© Printed in India

Connections for small vertex models

R SRINIVASAN

Mehta Research Institute of Mathematics and Mathematical Physics,
Allahabad 211 019, India

MS received 19 May 1999

Abstract. This paper is a first attempt at classifying connections on small vertex models i.e., commuting squares of the form displayed in (1.2) below. More precisely, if we let $B(k,n)$ denote the collection of matrices W for which (1.2) is a commuting square then, we: (i) obtain a simple model form for a representative from each equivalence class in $B(2,n)$, (ii) obtain necessary conditions for two such 'model connections' in $B(2,n)$ to be themselves equivalent, (iii) show that $B(2,n)$ contains a $(3n-6)$-parameter family of pairwise inequivalent connections, and (iv) show that the number $(3n-6)$ is sharp. Finally, we deduce that every graph that can arise as the principal graph of a finite depth subfactor of index 4 actually arises for one arising from a vertex model corresponding to $B(2,n)$ for some n.

Keywords. Subfactor; vertex model; biunitary.

1. Introduction

We first recall certain facts about commuting squares and biunitaries. These facts can be found in [USC] or [JS].

1.1 Consider the following commuting square:

$$\begin{tikzcd}
A_0^1 \arrow[r, phantom, "\overset{L}{\subset}"] & A_1^1 \\
A_0^0 \arrow[u, phantom, "K\cup"] \arrow[r, phantom, "\overset{G}{\subset}"] & A_1^0 \arrow[u, phantom, "\cup H"]
\end{tikzcd} \qquad (1.1)$$

Then the following are equivalent. (i) $G=L=[n]$ and $H=K=[k]$; (ii) the square (1.1) is isomorphic to a commuting square of the form

$$\begin{array}{ccc}
W(1\otimes M_k(\mathbb{C}))W^* & \subset & M_n(\mathbb{C})\otimes M_k(\mathbb{C}) \\
\cup & & \cup \\
\mathbb{C} & \subset & M_n(\mathbb{C})\otimes 1
\end{array}, \qquad (1.2)$$

where $W=((W_{\beta b}^{\alpha a}))\in M_n(\mathbb{C})\otimes M_k(\mathbb{C})$ is unitary. (We use the convention that $1\le\alpha$, $\beta\le n$, $1\le a$, $b\le k$.)

1.2 If $W=((W_{\beta b}^{\alpha a}))\in M_n(\mathbb{C})\otimes M_k(\mathbb{C})$, then the square (1.2) is a commuting square iff W is biunitary i.e., both W and $\tilde{W}$ given by $\tilde{W}_{\beta b}^{\alpha a}=W_{\alpha b}^{\beta a}$ are unitary.

We shall use the symbol $B(k,n)$ to denote the set of such biunitary matrices.

1.3 Two biunitary matrices W and W' are said to be equivalent if the corresponding commuting squares are isomorphic. It is true that if $W, W'\in B(k,n)$, then W and W' are equivalent if and only if there exists unitary matrices $U, U' \in M_n$, $A, A' \in M_k$ such that $(U \otimes A)W = W'(U' \otimes A')$.

1.4 Given $W \in B(k,n)$, the basic construction yields a grid of commuting squares and consequently, a *horizontal* (respectively *vertical*) subfactor $A_{\infty}^{0} \subset A_{\infty}^{1}$ (respectively $A_0^{\infty} \subset A_1^{\infty}$) with index k^2 (respectively n^2). This construction is canonical, and so, isomorphic commuting squares (i.e., equivalent biunitary matrices) yield isomorphic horizontal (respectively vertical) subfactors.

1.5 When $n = k = 2$, any $W \in B(2,2)$ is equivalent to a biunitary matrix of the form

$$W(\omega) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \omega \end{pmatrix},$$

where $\omega \in \mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$. Further neither the vertical nor the horizontal subfactor is irreducible. It is in fact true, although not mentioned in [USC], that $W(\omega)$ is equivalent to $W(\omega')$ if and only if $\mathrm{Re}(\omega) = \mathrm{Re}(\omega')$.

2. A model form for a matrix in $B(2,n)$

In this section we prove that every biunitary matrix in $B(2,n)$ is equivalent to a biunitary matrix in a model form with $(3n-5)$ independent parameters.

PROPOSITION 2.1

Any biunitary matrix $W \in B(2,n)$ is equivalent to a matrix of the form

$$\begin{pmatrix} C & US \\ VS & -UVC \end{pmatrix},$$

where U, V are diagonal unitary matrices and C, S are positive diagonal matrices such that $C^2 + S^2 = 1$.

Proof. Let $W = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, so that $\tilde{W} = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$ where $a, b, c, d \in M_n$. Then W is a biunitary matrix if and only if both W and $\tilde{W}$ are unitary matrices. The unitarity of W and $\tilde{W}$ (i.e. the relation $WW^* = 1 = \tilde{W}\tilde{W}^*$) implies the following equations:

$$\begin{aligned} aa^* + bb^* &= 1, & cc^* + dd^* &= 1 \\ a^*a + c^*c &= 1, & b^*b + d^*d &= 1 \\ aa^* + cc^* &= 1, & bb^* + dd^* &= 1 \\ a^*a + b^*b &= 1, & c^*c + d^*d &= 1. \end{aligned}$$

By premultiplying by $u \otimes 1$, where $u \in M_n$ is a suitable unitary matrix (i.e. by working with an equivalent biunitary matrix), we may assume, without loss of generality, that a is positive. Then it follows from the above equations that $0 \leq a \leq 1$. Let $C = a$. So there exist a unique positive matrix $S \in M_n$ such that, $0 \leq S \leq 1$ and $C^2 + S^2 = 1$. Then from the above equations we can conclude that b, c, d are normal and also that $bb^* = cc^* = S^2$ and $dd^* = C^2$. So there exist unitary matrices $U, V, T \in M_n$ such that $b = US, c = VS$ and $d = TC$, and such that U and V commute with S (hence, also with C) and T commutes with C (hence, also with S).

So we find that W is equivalent to the biunitary matrix

$$\begin{pmatrix} C & US \\ VS & TC \end{pmatrix},$$

where C, S, U, V and T are as above.

The biunitarity of W (i.e. the relation $WW^* = \tilde{W}\tilde{W}^* = 1$) also implies the following equations:

$$\begin{aligned} SC(V+TU^*) &= 0, \qquad & SC(U+V^*T) &= 0, \\ SC(U+TV^*) &= 0, \qquad & SC(V+U^*T) &= 0. \end{aligned} \tag{2.3}$$

Since U, V and T leave the eigenspaces $\{H_i\}_{i\in I}$ of C invariant, we may, by conjugating W by a unitary matrix of the form $\Gamma \otimes 1$ (where Γ is a unitary matrix which diagonalises C), assume that

$$\begin{aligned} &C = \oplus_{i\in I} c_i 1_{H_i}, \quad S = \oplus_{i\in I} s_i 1_{H_i} \\ &U = \oplus_{i\in I} U_i, \quad V = \oplus_{i\in I} V_i, \quad T = \oplus_{i\in I} T_i, \end{aligned}$$

where 1_{H_i} denotes the identity in $\mathcal{L}(H_i), 0 \leq c_i, s_i \leq 1$ and $U_i, V_i, T_i \in \mathcal{L}(H_i)$.

Thus we see that $W = \oplus_{i\in I} W_i$, where W_i is a biunitary matrix in $M_{n_i} \otimes M_2$ and n_i is the dimension of H_i, and that

$$W_i = \begin{pmatrix} c_i I_{n_i} & U_i s_i \\ s_i V_i & c_i T_i \end{pmatrix}.$$

Note that in order to complete the proof of this proposition it is enough if we prove that each of this W_i is equivalent to a biunitary matrix of the form presented in the proposition by pre- and post-multiplying by unitary matrices of the form $u_i \otimes 1$ (then by pre- and post-multiplying W by matrices of the form $(\oplus_{i\in I} u_i) \otimes 1$ one may prove that W is equivalent to a matrix of the required form). To prove this we now consider two cases depending on whether c_t is zero or non-zero.

Suppose $c_i = 0$, then, by pre-multiplying W_i by the unitary matrix $U_i^* \otimes 1$, we may assume that $U_i = 1_{H_i}$. Now, by conjugating W_i by a unitary matrix $p \otimes 1$, where p is a unitary matrix which diagonalises V_i, we can conclude that W_i is equivalent to a matrix of the desired form.

Suppose that $c_i \neq 0$. If $s_i = 0$ we can assume, by conjugating by a unitary matrix $p \otimes 1$ (where p is a unitary matrix which diagonalises T_i), that the matrix W_i is in the required form. If s_i is also non-zero, then first conclude from the set of equations (2.3) that $T_i = -U_i V_i = -V_i U_i$. Now, by conjugating W_i by a unitary matrix of the form $p \otimes 1$, where p is a unitary matrix which simultaneously diagonalises the commuting unitaries U_i, V_i, T_i, we may conclude that W_i is equivalent to a matrix of the desired form.

Hence, in any case, we find that W is equivalent to a matrix of the form

$$\begin{pmatrix} C & US \\ VS & -UVC \end{pmatrix},$$

where U, V are diagonal unitary matrices and C, S are positive diagonal matrices such that $C^2 + S^2 = 1$. □

It is proved in [USC] that when $n = k = 2$, any $W \in B(2,2)$ is equivalent to a biunitary matrix of the form

$$W(\omega)=\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&\omega\end{pmatrix},$$

where $\omega\in\mathbb{T}=\{z\in\mathbb{C}:|z|=1\}$. Further neither the vertical nor the horizontal subfactor is irreducible. We explicitly point out the ambiguity in such a representation.

PROPOSITION 2.2

$W(\omega)$ is equivalent to $W(\omega')$ if and only if $\mathrm{Re}(\omega)=\mathrm{Re}(\omega')$.

Proof. Let

$$U=\begin{pmatrix}0&1\\1&0\end{pmatrix},\quad A'=\begin{pmatrix}1&0\\0&\omega\end{pmatrix},\quad \text{and}\quad A=\begin{pmatrix}1&0\\0&1\end{pmatrix}.$$

Then it can be easily verified that $(U\otimes A)W(\omega)=W(\bar{\omega})(U\otimes A')$.

Conversely suppose

$$(U\otimes A)\begin{pmatrix}I&0\\0&D\end{pmatrix}=\begin{pmatrix}I&0\\0&D'\end{pmatrix}(U'\otimes A'),$$

where $D=\begin{pmatrix}1&0\\0&\omega\end{pmatrix}, D'=\begin{pmatrix}1&0\\0&\omega'\end{pmatrix}, A=\begin{pmatrix}a&b\\c&d\end{pmatrix}, A'=\begin{pmatrix}a'&b'\\c'&d'\end{pmatrix}$. Then the following equations hold:

$$aU=a'U', \tag{2.4}$$

$$bUD=b'U', \tag{2.5}$$

$$cU=c'D'U', \tag{2.6}$$

$$dUD=d'D'U'. \tag{2.7}$$

Suppose $a\neq 0$. Then $a'\neq 0$ by eq. (2.4), and so also $d'\neq 0$ (as D, D', U, U' are unitary matrices). From eqs (2.4) and (2.7), we see that $d'^{-1}da^{-1}a'U'DU'^{*}=D'$. Now by comparing the eigenvalues (as U' is an unitary matrix), we conclude that $\{d'^{-1}da^{-1}a', d'^{-1}da^{-1}a'\omega\}=\{1,\omega'\}$. Hence either $\omega=\omega'$ or $\omega=\bar{\omega}'$. Suppose that $a=0$, then as A is an unitary matrix, it is the case that $b\neq 0$. Exactly in a similar way, using eqs (2.5) and (2.6), we can again conclude that either $\omega=\omega'$ or $\omega=\bar{\omega}'$. □

PROPOSITION 2.3

Any $W\in B(2,n)$ is equivalent to a biunitary matrix of the form

$$W(\omega,\theta,\phi,C)=\begin{pmatrix}1&0&0&0&0&0\\0&1&0&0&0&0\\0&0&C&0&0&\theta S\\0&0&0&1&0&0\\0&0&0&0&\omega&0\\0&0&\phi S&0&0&-\theta\phi C\end{pmatrix},$$

where $\theta=\mathrm{diag}(\theta_1,\theta_2,\ldots,\theta_{n-2})$, $\phi=\mathrm{diag}(\phi_1,\phi_2,\ldots,\phi_{n-2})$, $C=\mathrm{diag}(C_1,C_2,\ldots,C_{n-2})$, $S=\mathrm{diag}(S_1,S_2,\ldots,S_{n-2})$, $\theta_i,\phi_i,\omega\in\{z\in\mathbb{C}:|z|=1\}$, $\mathrm{Im}(\omega)\geq 0, 0\leq C_i,S_i\leq 1$, *and* $C_i^2+S_i^2=1$.

Proof. From Proposition 2.1, we may assume that $W = \sum_{i=1}^{n} E_{ii} \otimes W_i$ (where W_i is a 2×2 unitary matrix and $\{E_{ij} : 1 \leq i, j \leq n\}$ denotes – here and elsewhere – the usual system of matrix units in M_n), and that W_i has the form

$$W_i = \begin{pmatrix} C_i & \theta_i S_i \\ \phi_i S_i & -\theta_i \phi_i C_i \end{pmatrix},$$

where θ_i, ϕ_i are complex numbers of unit modulus, and $0 \leq C_i, S_i \leq 1$ and $C_i^2 + S_i^2 = 1$.

Note next that if $D = \text{diag}(d_1, \ldots, d_n) \in M_n$ is a diagonal unitary matrix, and if W, W_i are as above, and if $V_1, V_2 \in M_2$ are unitary, then

$$(D \otimes V_1)W(1 \otimes V_2) = \sum_{i=1}^{n} d_i(E_{ii} \otimes V_1 W_i V_2). \tag{$*$}$$

Set $V_1 = 1, V_2 = W_1^*$; if the $(1,1)$ entry of $W_i W_1^*$ is $\omega_i \tilde{C}_i$, with $\tilde{C}_i \geq 0$ and $|\omega_i| = 1$, define $d_i = \bar{\omega}_i$. We may now deduce from equation $(*)$ that we may reduce to the case where W_1 is the identity matrix, and W_i are as above.

Next, let U be the unitary matrix which diagonalises (the new) W_2. Then, by setting $d_i = \overline{\omega_i}$ if $\omega_i \tilde{C}_i'$ is the $(1,1)$ entry of $U^* W_i U$, with $\tilde{C}_i' \geq 0$ and $|\omega_i| = 1$, and by setting $U = V_1^* = V_2$, we find that we may reduce to the case where W is as above, and in addition, $W_1 = 1$ and $W_2 = \text{diag}(1, \omega)$, where ω is a complex number of unit modulus.

If $\text{Im}(\omega) \geq 0$, the proof of the Proposition is complete. If $\text{Im}(\omega) < 0$, then set

$$V_1 = V_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

and $d_1 = 1$, $d_2 = \bar{\omega}, d_i = -\overline{\theta_i \phi_i}\, \forall i = 3, \ldots, n$, to conclude that W is indeed equivalent to a biunitary matrix of the prescribed form. □

3. Classification of $B(2,n)$

We shall use the notation $\Omega(2,n) = \mathbb{T}^+ \times \mathbb{T}^{n-2} \times \mathbb{T}^{n-2} \times [0,1]^{n-2}$, where $\mathbb{T}$ is the unit circle in the complex plane and $\mathbb{T}^+ = \{\omega \in \mathbb{T} : \text{Im}(\omega) \geq 0\}$; we shall denote a typical pair of points of $\Omega(2,n)$ by $P = (\omega, \theta, \phi, C)$ and $P' = (\omega', \theta', \phi', C')$ and the corresponding biunitary matrices by W and W'. Also we shall denote the matrices $\begin{pmatrix} 1 & 0 \\ 0 & \omega \end{pmatrix}$ and $\begin{pmatrix} 1 & 0 \\ 0 & \omega' \end{pmatrix}$ by D and D' respectively.

We isolate a simple assertion as a lemma, since we will need to repeatedly use it.

Lemma 3.1. *Suppose a, b, c, d are non-zero complex numbers, and suppose θ, ϕ, ω_j, $j = 0, 1, 2$ are complex numbers of unit modulus, and suppose C and S are non-negative real numbers satisfying $C^2 + S^2 = 1$. Assume that* $\text{Im}(\omega_2) > 0$ *and that the following equations are satisfied:*

$$a(C - \omega_0) + b\phi S = 0, \tag{3.8}$$

$$a\theta S - b(\theta\phi C + \omega_0\omega_1) = 0, \tag{3.9}$$

$$c(C - \omega_0\omega_2) + d\phi S = 0, \tag{3.10}$$

$$c\theta S - d(\theta\phi C + \omega_0\omega_1\omega_2) = 0. \tag{3.11}$$

Then, $S \neq 0, C \neq 1$ *and*

$$\mathrm{Re}(\omega_2) = -[S^2 + \mathrm{Re}(\bar{\omega}_1\theta\phi)C^2]. \tag{3.12}$$

Proof. If $S = 0$, then the eqs (3.8) and (3.10) would imply that $\omega_0 = \omega_0\omega_2 = 1$. Hence, as we have assumed that $\omega_2 \neq 1$, conclude that $S \neq 0$. Now, deduce from (3.8) and (3.9) that

$$(\omega_0 - C)(\theta\phi C + \omega_0\omega_1) = \theta\phi S^2; \tag{3.13}$$

similarly, deduce from eqs (3.10) and (3.11) that

$$(\omega_0\omega_2 - C)(\theta\phi C + \omega_0\omega_1\omega_2) = \theta\phi S^2.$$

These equations may be re-written as

$$\begin{aligned}&\omega_0^2\omega_1 + \omega_0(\theta\phi - \omega_1)C = \theta\phi S^2,\\ &\omega_0^2\omega_2^2\omega_1 + \omega_0\omega_2(\theta\phi - \omega_1)C = \theta\phi S^2,\end{aligned}$$

from which we may deduce that

$$\omega_0^2(1 - \omega_2^2)\omega_1 + \omega_0(1 - \omega_2)(\theta\phi - \omega_1)C = 0. \tag{3.14}$$

As we have assumed that $\mathrm{Im}(\omega_2) > 0$, we may infer from eq. (3.14) that

$$\omega_0 = \frac{(\omega_1 - \theta\phi)C}{\omega_1(1 + \omega_2)}.$$

Substituting this expression for ω_0 into eq. (3.13), we find that

$$\begin{aligned}(\omega_0 - C)(\theta\phi C + \omega_0\omega_1) &= \frac{C}{\omega_1(1 + \omega_2)}(\omega_1 - \theta\phi - \omega_1(1 + \omega_2))\\ &\quad\times \frac{C}{\omega_1(1 + \omega_2)}(\theta\phi\omega_1(1 + \omega_2) + \omega_1(\omega_1 - \theta\phi))\\ &= -\frac{C^2}{(1 + \omega_2)^2}(\bar{\omega}_1\theta\phi + \omega_2)(\omega_1 + \theta\phi\omega_2);\end{aligned}$$

and hence,

$$\theta\phi S^2 - (\omega_0 - C)(\theta\phi C + \omega_0\omega_1) = \theta\phi\left[S^2 + \frac{C^2}{(1 + \omega_2)^2}(\bar{\omega}_1\theta\phi + \omega_2)(\omega_1\overline{\theta\phi} + \omega_2)\right].$$

Thus we find that the equation $(\omega_0 - C)(\theta\phi C + \omega_0\omega_1) = \theta\phi S^2$ will be satisfied precisely when

$$\begin{aligned}0 &= (1 + \omega_2)^2 S^2 + C^2(\bar{\omega}_1\theta\phi + \omega_2)(\omega_1\overline{\theta\phi} + \omega_2)\\ &= \omega_2^2(S^2 + C^2) + \omega_2(2S^2 + 2C^2\,\mathrm{Re}(\bar{\omega}_1\theta\phi)) + (S^2 + C^2)\\ &= \omega_2^2 - 2\alpha\omega_2 + 1, (\text{say}),\end{aligned}$$

where $\alpha = -(S^2 + C^2\mathrm{Re}(\bar{\omega}_1\theta\phi))$.

On the other hand, it is clear that if a complex number ω satisfies the equation $\omega^2 - 2\alpha\omega + 1 = 0$, where α is real and $|\alpha| \leq 1$, then $\omega = \alpha \pm i\sqrt{1 - \alpha^2}$, so that $\mathrm{Re}(\omega) = \alpha$, and hence eq. (3.12) is satisfied. □

In the next proposition we give a partial classification of $B(2,n)$. $B(2,3)$ is completely classified in [Sr].

PROPOSITION 3.2

Let n be arbitrary. Assume that $\mathrm{Im}(\omega)$, $\mathrm{Im}(\omega') > 0$, *and* $C_i, S_i, C_i', S_i' \neq 0$ *for all i.*

(a) *If* $W(\omega,\theta,\phi,C)$ *is equivalent to* $W(\omega',\theta',\phi',C')$, *then one of the following relations holds:*

(0) $\omega = \omega'$, *and there exists a permutation* $\sigma \in S_{n-2}$ *such that*

$$\mathrm{Ad}(P_\sigma)(C) = C', \quad \mathrm{Ad}(P_\sigma)(\theta) = \zeta\theta', \quad \textit{and} \quad \mathrm{Ad}(P_\sigma)(\phi) = \bar{\zeta}\phi',$$

where ζ *is some complex number of unit modulus,* P_σ *denotes the permutation matrix corresponding to* σ, *and we write* $\mathrm{Ad}(P_\sigma) = P_\sigma(\cdot)P_\sigma^{-1}$.

(1) $\omega = \omega'$, *and there exists a permutation* $\sigma \in S_{n-2}$ *such that*

$$\mathrm{Ad}(P_\sigma)(C) = C', \quad \mathrm{Ad}(P_\sigma)(\theta) = \zeta\theta'^*, \quad \textit{and} \quad \mathrm{Ad}(P_\sigma)(\phi) = \omega\bar{\zeta}\phi'^*,$$

where ζ *is some complex number of unit modulus.*

(2) *There exist i, i' such that* $(\mathrm{Re}(\omega'),\ \mathrm{Re}(\omega)) \in \Lambda_i \times \Lambda_{i'}'$, *where* $\Lambda_i = \{-(S_i^2 + \mathrm{Re}(\theta_i\phi_i)C_i^2),\ -(S_i^2 + \mathrm{Re}(\bar{\omega}\,\theta_i\phi_i)C_i^2)\}$ *and* $\Lambda_i' = \{-(S_i'^2 + \mathrm{Re}(\theta_i'\phi_i')C_i'^2),\ -(S_i'^2 + \mathrm{Re}(\bar{\omega}'\theta_i'\phi_i')C_i'^2)\}$.

(3) *There exist* $i,\ j, i', j'$ *such that* $(\mathrm{Re}(\omega'),\ \mathrm{Re}(\omega)) = (-m_{i,j}, -m_{i',j'}')$ *where* $m_{i,j} = 1 - (1 + \mathrm{Re}(\theta_i\phi_i\bar{\theta}_j\bar{\phi}_j))C_i^2C_j^2 - (1 + \mathrm{Re}(\theta_i\phi_j\bar{\theta}_j\bar{\phi}_i))S_i^2S_j^2 - 2(\mathrm{Re}(\theta_i\bar{\theta}_j) + \mathrm{Re}(\phi_i\bar{\phi}_j))C_iC_jS_iS_j$, *and* $m_{i,j}'$ *is the corresponding 'primed' expression.*

(b) *In* (a), *conditions* (0) *and* (1) *are also sufficient conditions for* $W(\omega,\theta,\phi,C)$ *to be equivalent to* $W(\omega,\theta',\phi',C')$.

(c) (i) *The vertical subfactor associated with* $W(\omega,\theta,\phi,C)$ *is always reducible.*

(ii) *The horizontal subfactor associated with* $W(\omega,\theta,\phi,C)$ *is reducible if and only if either of the following two conditions holds:*

(1) $S = 0$;

(2) $\omega = 1$ *and there exists scalars* $\lambda_1 \in \mathbb{T}$ *and* $\lambda_2 \in \mathbb{C}$ *such that* $\phi S = \lambda_1\theta S$ *and* $(1+\theta\phi)C = \lambda_2\theta S$.

Proof. First we write the condition for $P, P' \in \Omega(2,n)$ to afford equivalent connections, as a set of equations. Thus, in order for W to be equivalent to W', i.e. $(U \otimes A)W(\omega,\theta,\phi,C) = W(\omega',\theta',\phi',C')(U' \otimes A')$, where

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad A' = \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} \in U(2),$$

$$U = \begin{pmatrix} u_{1,1} & u_{1,2} & P^t \\ u_{2,1} & u_{2,2} & Q^t \\ X & Y & Z \end{pmatrix}, \quad U' = \begin{pmatrix} u_{1,1}' & u_{1,2}' & P'^t \\ u_{2,1}' & u_{2,2}' & Q'^t \\ X' & Y' & Z' \end{pmatrix} \in U(n),$$

$P, Q, X, Y, P', Q', X', Y' \in M_{(n-2)\times 1}$ and $Z, Z' \in M_{n-2}$, where X^t denotes the matrix transpose of X, it is necessary and sufficient that the following set of equations holds.

$$u_{1,1}A = u_{1,1}'A', \tag{3.15}$$

$$u_{1,2}AD = u_{1,2}'A', \tag{3.16}$$

$$P^t(aC + b\phi S) = d'P'^t, \tag{3.17}$$

$$P^t(a\theta S - b\theta\phi C) = b'P'^t, \qquad (3.18)$$

$$P^t(cC + d\phi S) = c'P'^t, \qquad (3.19)$$

$$P^t(c\theta S - d\theta\phi C) = d'P'^t, \qquad (3.20)$$

$$u_{2,1}A = u'_{2,1}D'A', \qquad (3.21)$$

$$u_{2,2}AD = u'_{2,2}D'A', \qquad (3.22)$$

$$Q^t(aC + b\phi S) = a'Q'^t, \qquad (3.23)$$

$$Q^t(a\theta S - b\theta\phi C) = b'Q'^t, \qquad (3.24)$$

$$Q^t(cC + d\phi S) = c'\omega'Q'^t, \qquad (3.25)$$

$$Q^t(c\theta S - d\theta\phi C) = d'\omega'^t Q'^t, \qquad (3.26)$$

$$aX = (a'C' + c'\theta'S')X', \qquad (3.27$$

$$bX = (b'C' + d'\theta'S')X', \qquad (3.28$$

$$cX = (a'\phi'S' - c'\theta'\phi'C')X', \qquad (3.29$$

$$dX = (b'\phi'S' - d'\theta'\phi'C')X', \qquad (3.30$$

$$aY = (a'C' + c'\theta'S')Y', \qquad (3.31$$

$$b\omega Y = (b'C' + d'\theta'S')Y', \qquad (3.32$$

$$cY = (a'\phi'S' - c'\theta'\phi'C')Y', \qquad (3.33$$

$$d\omega Y = (b'\phi'S' - d'\theta'\phi'C')Y', \qquad (3.34$$

$$Z(aC + b\phi S) = (a'C' + c'\theta'S')Z', \qquad (3.35$$

$$Z(a\theta S - b\theta\phi C) = (b'C' + d'\theta'S')Z', \qquad (3.36$$

$$Z(cC + d\phi S) = (a'\phi'S' - c'\theta'\phi'C')Z', \qquad (3.37$$

$$Z(c\theta S - d\theta\phi C) = (b'\phi'S' - d'\theta'\phi'C')Z'. \qquad (3.38$$

Now we consider cases depending on whether various entries of U are zero or non zero.

Case (1): $u_{1,1} \neq 0$. The unitarity of A and A', together with eq. (3.15) imply that $|u_{1,1}| =$ $|u'_{1,1}|$. Let $u_{1,1} = zu'_{1,1}$ where $|z| = 1$; it follows that $A' = zA$. So (by replacing the pai (A, U) by $(zA, z^{-1}U)$, in case $z \neq 1$) we may assume, without loss of generality, tha $A = A'$ and $u_{1,1} = u'_{1,1}$.

Since A is a unitary matrix deduce from (3.16) that $u_{1,2}D = u'_{1,2}I_2$, where I_2 denotes th identity matrix in M_2. The assumption $\omega \neq 1$ now implies that $u_{1,2} = u'_{1,2} = 0$.

Similarly eq. (3.21) implies that $u_{2,1}I_2 = D'u'_{2,1}$. Again the assumption that $\omega' \neq$ implies that $u_{2,1} = u'_{2,1} = 0$.

We consider two sub-cases depending upon whether the entry $u_{2,2}$ is non-zero c zero.

Case (1.1): $u_{2,2} \neq 0$. As $\mathrm{Im}(\omega)$, $\mathrm{Im}(\omega') > 0$. First we deduce from eq. (3.22) that eithe $a = 0$ or $b = 0$. But $a = d = 0$ implies that $\omega = \bar{\omega}'$. But as we have assumed th $\mathrm{Im}(\omega)$, $\mathrm{Im}(\omega') > 0$, it is the case that $b = c = 0$, and that $\omega = \omega'$, and $u_{2,2} = u'_{2,2}$. We wi show that the relation (0) is satisfied in this case.

As $S_i \neq 0$ for all i (by the assumption in the statement of the Proposition), we find fro eq. (3.18) and (3.24) that $P^t = Q^t = 0$. At the same time eqs (3.17) and (3.23) imply th $P'^t = Q'^t = 0$. Also by the assumption $S'_i \neq 0$ for all i, we find from eqs (3.28) and (3.3

that $X' = Y' = 0$, while the eqs (3.27) and (3.31) imply that $X = Y = 0$. The unitarity of U and U' is now seen to imply that Z and Z' also are unitary.

Equations (3.35) and (3.38) may be rewritten as

$$ZC = C'Z', \tag{3.39}$$

$$ZC\theta\phi = \theta'\phi'C'Z'. \tag{3.40}$$

Since C and C' are invertible positive operator (as follows from the assumption in the statement of the Proposition that $C_i \neq 0$ for all i) and since eq. (3.39) may be re-written as

$$(ZZ'^*)(Z'CZ'^*) = C',$$

we may deduce from the uniqueness of polar decomposition that $Z = Z'$.

Next, we may deduce from eqs (3.39) and (3.40) – using the invertibility of the matrix C' – that $Z\theta\phi Z^* = \theta'\phi'$.

Thus,

$$ZC = C'Z \quad \text{(hence also } ZS = S'Z\text{) and } Z\theta\phi = \theta'\phi'Z.$$

Notice now that

$$\begin{aligned} Z\theta S &= Z(\theta\phi)\phi^* S \\ &= (\theta'\phi')Z\phi^* S \\ &= (\theta'\phi')(Z\phi^* Z^*)ZS \\ &= (\theta'\phi')(Z\phi^* Z^*)S'Z. \end{aligned}$$

Hence, we may deduce from (3.36) that

$$a(\theta'\phi')(Z\phi^*Z^*)S'Z = d\theta'S'Z;$$

deduce from the invertibility of $S'Z$ that

$$Z\phi^*Z^* = \zeta\phi'^*,$$

where $\zeta = d/a$. Since $Z\theta\phi Z^* = (\theta'\phi')$, we thus find that

$$Z\theta Z^* = \zeta\theta'.$$

Let $\mathcal{A}$ (resp., $\mathcal{A}'$) denote the *-subalgebra of M_{n-2} generated by $\{\theta, \phi, C\}$ (resp., $\{\theta', \phi', C'\}$). The preceding analysis shows that the map $\mathrm{Ad}(Z) = Z(\cdot)Z^*$ maps $\mathcal{A}$ onto $\mathcal{A}'$ (since it carries the generators to non-zero multiples of the generators).

Note now that $\mathcal{A}$ and $\mathcal{A}'$ are contained in the algebra of diagonal matrices. If $\{e_\alpha : \alpha \in \Lambda\}$ denotes the set of minimal projections in the abelian C^*-algebra $\mathcal{A}$, and if $Ze_\alpha Z^* = e'_\alpha$, then clearly $\{e'_\alpha : \alpha \in \Lambda\}$ is the set of minimal projections in $\mathcal{A}'$. The fact that some unitary matrix i.e., Z – simultaneously conjugates each e_α into e'_α, clearly implies now that we can find some permutation $\sigma \in S_{n-2}$ such that $\mathrm{Ad}(P_\sigma)$ maps each e_α into e'_α.

It follows easily now from the construction that

$$\mathrm{Ad}(P_\sigma)(\theta) = \zeta\theta', \mathrm{Ad}(P_\sigma)(\phi) = \bar{\zeta}\phi', \quad \text{and} \quad \mathrm{Ad}(P_\sigma)(C) = C'.$$

Case (1.2): $u_{2,2} = 0$. We will prove that the relation (2) is satisfied in this case. From the eq. (3.22) we conclude that $u'_{2,2} = 0$. Suppose $Q^t = (q_1, q_2, \dots, q_{n-2})$. As $u_{2,1} = u_{2,2} = 0$, we find that Q is a unit vector; hence there exists an index i such that $q_i \neq 0$. Then, we find from eqs (3.23)–(3.26) that

$$\begin{aligned} q_i(aC_i + b\phi_i S_i) &= aq'_i, && (3.41)\\ q_i(a\theta_i S_i - b\theta_i\phi_i C_i) &= bq'_i, && (3.42)\\ q_i(cC_i + d\phi_i S_i) &= c\omega' q'_i, && (3.43)\\ q_i(c\theta_i S_i - d\theta_i\phi_i C_i) &= d\omega' q'_i. && (3.44)\end{aligned}$$

The unitarity of the matrix $\begin{pmatrix} C_i & \phi_i S_i \\ \theta_i S_i & -\theta_i\phi_i C_i \end{pmatrix}$ would imply necessarily that $|q'_i| = |q_i| \neq 0$. Also, as $S_i \neq 0$, we may infer from eqs (3.41) and (3.42) that $a,b,c,d \neq 0$. Let $Y^t = (y_1, y_2, \ldots, y_{n-2})$. Since $u_{1,2} = u_{2,2} = 0$, we find that Y is a unit vector; hence there exists an index i' such that $y_{i'} \neq 0$. Then, we find from eqs (3.31)–(3.34), that the following equations hold:

$$\begin{aligned} ay_{i'} &= (aC'_{i'} + c\theta'_{i'}S'_{i'})y'_{i'},\\ b\omega y_{i'} &= (bC'_{i'} + d\theta'_{i'}S'_{i'})y'_{i'},\\ cy_{i'} &= (a\phi'_{i'}S'_{i'} - c\theta'_{i'}\phi'_{i'}C'_{i'})y'_{i'},\\ d\omega y_{i'} &= (b\phi'_{i'}S'_{i'} - d\theta'_{i'}\phi'_{i'}C'_{i'})y'_{i'}. && (3.45)\end{aligned}$$

Again, using the unitarity of the matrix $\begin{pmatrix} C'_{i'} & \phi'_{i'}S'_{i'} \\ \theta'_{i'}S'_{i'} & -\theta'_{i'}\phi'_{i'}C'_{i'} \end{pmatrix}$, deduce that $y_{i'}$ and $y'_{i'}$ have the same absolute value.

Let $\omega_0 = q'_i q_i^{-1}$ and $\omega'_0 = y'_{i'} y_{i'}^{-1}$. Now, first by re-writing the above two equations in the form as in eqs (3.8)–(3.11), and then by applying Lemma 3.1 separately to the two sets of equations above conclude that

$$(\mathrm{Re}(\omega'), \mathrm{Re}(\omega)) = (-(S_i^2 + \mathrm{Re}(\theta_i\phi_i)C_i^2), -(S'^2_{i'} + \mathrm{Re}(\theta'_{i'}\phi'_{i'})C'^2_{i'})).$$

Hence the relation (2) is satisfied in this case.

Case (2): $u_{1,1} = 0$.

Case (2.1): $u_{1,2} \neq 0$. Using (3.16) and the unitarity of A and A', we can assume without loss of generality that $AD = A'$ and $u_{1,2} = u'_{1,2}$. Also as $\omega' \neq 1$, we find from (3.22) that $u_{2,2} = u'_{2,2} = 0$. There are two cases now, depending on whether $u_{2,1}$ is not or is 0, which we consider separately.

Case (2.1.1): $u_{2,1} \neq 0$. As $\mathrm{Im}(\omega), \mathrm{Im}(\omega') > 0$ we may deduce from (3.21) that $a = d = 0$, $\omega = \omega'$, and $u_{2,1} = \omega u'_{2,1}$. We will show that the relation (1) is satisfied in this case.

As S is invertible, (i.e. $S_i \neq 0$ for all i) we find from (3.17) and (3.23) that $P^t = Q^t = 0$. From (3.18) and (3.24), we get $P'^t = Q'^t = 0$. As S' is invertible, we find from (3.27) and (3.31) that $X' = Y' = 0$, and then from (3.28) and (3.32) we get $X = Y = 0$. Now it follows that Z and Z' are unitary.

From (3.36) and (3.37) we have

$$\begin{aligned} -Z\theta\phi C &= \omega C'Z',\\ ZC &= -\theta'\phi'C'Z'. \end{aligned}$$

It follows (as before, from the uniqueness of polar decomposition and the invertibility of the positive operators C, C') that $Z\theta\phi = -\omega Z'$ and $Z = -\theta'\phi'Z'$. These equations together with the equation $bZ\phi S = c\theta'S'Z'$ (which is a consequence of (3.35)) are seen to imply (after some minor manipulations) that

$$\mathrm{Ad}(Z')(C) = C', \quad \mathrm{Ad}(Z')(\theta) = \zeta\theta'^{*}, \quad \text{and} \quad \mathrm{Ad}(Z')(\phi) = \omega\bar{\zeta}\phi'^{*},$$

where ζ is some scalar of unit modulus.

Arguing exactly as in the proof of Case (1.1), we may deduce the existence of a permutation $\sigma \in S_{n-2}$ such that

$$\mathrm{Ad}(P_\sigma)(C) = C', \quad \mathrm{Ad}(P_\sigma)(\theta) = \zeta\theta'^{*}, \quad \text{and} \quad \mathrm{Ad}(P_\sigma)(\phi) = \omega\bar{\zeta}\phi'^{*}.$$

Case (2.1.2): $u_{2,1} = 0$. It follows from (3.21) that $u'_{2,1} = 0$. Using the unitarity of U and the fact that $(u_{2,1}, u_{2,2}) = 0$, deduce that $Q^t(= (q_1, \ldots, q_{n-2}))$ is a unit vector and hence there exists an index i such that $q_i \neq 0$. Similarly the unitarity of U and the fact that $(u_{1,1}, u_{1,2}) = 0$, implies that the vector $X(= (x_1, \ldots, x_{n-2}))$ is a unit vector and hence there exists an index i' such that $x_{i'} \neq 0$.

Then, we find from eqs (3.23)–(3.26) that

$$\begin{aligned} q_i(aC_i + b\phi_i S_i) &= aq'_i, \\ q_i(a\theta_i S_i - b\theta_i\phi_i C_i) &= b\omega q'_i, \\ q_i(cC_i + d\phi_i S_i) &= c\omega' q'_i, \\ q_i(c\theta_i S_i - d\theta_i\phi_i C_i) &= d\omega\omega' q'_i. \end{aligned} \tag{3.46}$$

Also we find from (3.27)–(3.30) that the following equations hold:

$$\begin{aligned} ax_{i'} &= (aC'_{i'} + c\theta'_{i'}S'_{i'})x'_{i'}, \\ b\bar{\omega}x_{i'} &= (bC'_{i'} + d\theta'_{i'}S'_{i'})x'_{i'}, \\ cx_{i'} &= (a\phi'_{i'}S'_{i'} - c\theta'_{i'}\phi'_{i'}C'_{i'})x'_{i'}, \\ d\bar{\omega}x_{i'} &= (b\phi'_{i'}S'_{i'} - d\theta'_{i'}\phi'_{i'}C'_{i'})x'_{i'}. \end{aligned}$$

Now using the unitarity of the matrix $\begin{pmatrix} C_i & \phi_i S_i \\ \theta_i S_i & -\theta_i\phi_i C_i \end{pmatrix}$ $\left(\text{resp. the matrix } \begin{pmatrix} C'_{i'} & \phi'_{i'}S'_{i'} \\ \theta'_{i'}S'_{i'} & -\theta'_{i'}\phi'_{i'}C'_{i'} \end{pmatrix}\right)$ deduce that $|q'_i| = |q_i| \neq 0$ (resp. $|x_i| = |x'_i|$). Also, as $S_i \neq 0$, we may infer from the set of equations (3.46) that $a, b, c, d \neq 0$.

Let $\omega_0 = q'_i q_i^{-1}$ and $\omega'_0 = x'_{i'} x_{i'}^{-1}$. Now, by applying Lemma 3.1 twice to the two sets of equations above (exactly as before), conclude that

$$(\mathrm{Re}(\omega'), \mathrm{Re}(\omega)) = (-(S_i^2 + \mathrm{Re}(\bar{\omega}\theta_i\phi_i)C_i^2), -(S_{i'}'^2 + \mathrm{Re}(\theta'_{i'}\phi'_{i'})C_{i'}'^2)).$$

Hence the relation (2) is satisfied in this case.

Case (2.2): $u_{1,2} = 0$. We break this into cases depending on whether $u_{2,1}$ vanishes or not.

Case (2.2.1): $u_{2,1} \neq 0$. As before using the unitarity of A, A' and (3.21) we may assume that $u_{2,1} = u'_{2,1}$ and $A = D'A'$. Using the unitarity of U and the fact that $(u_{1,1}, u_{1,2}) = 0$, deduce that $P^t(= (p_1, \ldots, p_{n-2}))$ is a unit vector and hence that there exists an index i such that $p_i \neq 0$. As $\omega \neq 1$, the matrix D is linearly independent from the identity matrix. Hence using (3.22) conclude that $u_{2,2} = u_{2,2} = 0$. Now the unitarity of U and the fact that $(u_{1,2}, u_{2,2}) = 0$, implies that the vector $Y(= (y_1, \ldots, y_{n-2}))$ is a unit vector and hence that there exists an index i' such that $y_{i'} \neq 0$.

Then, we find from (3.17)–(3.20) that

$$\begin{aligned} p_i(aC_i + b\phi_i S_i) &= ap'_i, \\ p_i(a\theta_i S_i) - b\theta_i\phi_i C_i) &= bp'_i, \\ p_i(cC_i + d\phi_i S_i) &= c\bar{\omega}' p'_i, \\ p_i(c\theta_i S_i - d\theta_i\phi_i C_i) &= d\bar{\omega}' p'_i. \end{aligned}$$

Also we find from the (3.31)–(3.34) that the following equations hold:

$$\begin{aligned} ay_{i'} &= (aC'_{i'} + c\bar{\omega}'\theta'_{i'}S'_{i'})y'_{i'}, \\ b\omega y_{i'} &= (bC'_{i'} + d\bar{\omega}'\theta'_{i'}S'_{i'})y'_{i'}, \\ cy_{i'} &= (a\phi'_{i'}S'_{i'} - c\bar{\omega}'\theta'_{i'}\phi'_{i'}C'_{i'})y'_{i'}, \\ d\omega y_{i'} &= (b\phi'_{i'}S'_{i'} - d\bar{\omega}'\theta'_{i'}\phi'_{i'}C'_{i'})y'_{i'}. \end{aligned}$$

Again the unitarity of the matrix $\begin{pmatrix} C_i & \phi_i S_i \\ \theta_i S_i & -\theta_i\phi_i C_i \end{pmatrix}$ $\left(\text{resp. the matrix } \begin{pmatrix} C'_{i'} & \phi'_{i'}S'_{i'} \\ \theta'_{i'}S'_{i'} & -\theta'_{i'}\phi'_{i'}C'_{i'} \end{pmatrix}\right)$ implies that $|p'_i| = |p_i| \neq 0$ (resp. $|y_i| = |y'_i|$). Also, as $S_i \neq 0$, we may infer from the above set of equations that $a, b, c, d \neq 0$.

Let $\omega_0 = p'_i p_i^{-1}$ and $\omega'_0 = y'_{i'} y_{i'}^{-1}$. Now, by applying Lemma 3.1 twice to the two sets of equations above (exactly as before), conclude that

$$(\mathrm{Re}(\omega'),\ \mathrm{Re}(\omega)) = (-(S_i^2 + \mathrm{Re}(\theta_i\phi_i)C_i^2), -(S_{i'}^{\prime 2} + \mathrm{Re}(\bar{\omega}'\theta'_{i'}\phi'_{i'})C_{i'}^{\prime 2})).$$

Hence the relation (2) is satisfied in this case.

Case (2.2.2): $u_{2,1} = 0$. First, suppose $u_{2,2} \neq 0$. Then, using the unitarity of A and equation (3.22), we may assume without loss of generality that $u_{2,2} = u'_{2,2}$ and $AD = D'A'$. As before using the unitarity of U and the fact that $(u_{1,1}, u_{1,2}) = 0$, deduce that $P^t(= (p_1, \dots, p_{n-2}))$ is a unit vector and hence there exists an index i such that $p_i \neq 0$. Also the unitarity of U and the fact that $(u_{1,1}, u_{2,1}) = 0$, implies that the vector $X(= (x_1, \dots, x_{n-2}))$ is a unit vector and hence that there exists an index i' such that $x_{i'} \neq 0$.

Then, we find from eqs (3.17)–(3.20) that

$$\begin{aligned} p_i(aC_i + b\phi_i S_i) &= ap'_i, \\ p_i(a\theta_i S_i - b\theta_i\phi_i C_i) &= b\omega p'_i, \\ p_i(cC_i + d\phi_i S_i) &= c\bar{\omega}' p'_i, \\ p_i(c\theta_i S_i - d\theta_i\phi_i C_i) &= d\omega\bar{\omega}' p'_i. \end{aligned}$$

Also we find from (3.27)–(3.30) that the following equations hold:

$$\begin{aligned} ax_{i'} &= (aC'_{i'} + c\bar{\omega}'\theta'_{i'}S'_{i'})x'_{i'}, \\ b\bar{\omega} x_{i'} &= (bC'_{i'} + d\bar{\omega}'\theta'_{i'}S'_{i'})x'_{i'}, \\ cx_{i'} &= (a\phi'_{i'}S'_{i'} - c\bar{\omega}'\theta'_{i'}\phi'_{i'}C'_{i'})x'_{i'}, \\ d\bar{\omega} x_{i'} &= (b\phi'_{i'}S'_{i'} - d\omega'\theta'_{i'}\phi'_{i'}C'_{i'})x'_{i'}. \end{aligned}$$

Again the unitarity of the matrix $\begin{pmatrix} C_i & \phi_i S_i \\ \theta_i S_i & -\theta_i\phi_i C_i \end{pmatrix}$ $\left(\text{resp. the matrix } \begin{pmatrix} C'_{i'} & \phi'_{i'}S'_{i'} \\ \theta'_{i'}S'_{i'} & -\theta'_{i'}\phi'_{i'}C'_{i'} \end{pmatrix}\right)$ implies that $|p'_i| = |p_i| \neq 0$ (resp. $|x_i| = |x'_i|$). Also, as $S_i \neq 0$, we may infer from the abov set of equations that $a, b, c, d \neq 0$.

Let $\omega_0 = p'_i p_i^{-1}$ and $\omega'_0 = x'_{i'} x_{i'}^{-1}$. Now, by applying Lemma 3.1 twice to the two sets of equations above, conclude exactly as before that

$$(\mathrm{Re}(\omega'),\ \mathrm{Re}(\omega)) = (-(S_i^2 + \mathrm{Re}(\bar{\omega}\theta_i\phi_i)C_i^2), -(S_{i'}'^2 + \mathrm{Re}(\bar{\omega}'\theta'_{i'}\phi'_{i'})C_{i'}'^2)).$$

Hence the relation (2) is satisfied in this case also.

Next we consider the final case when $u_{2,2}$ is also zero.

Case 2.2.3: $\begin{pmatrix} u_{1,1} & u_{1,2} \\ u_{2,1} & u_{2,2} \end{pmatrix} = 0$. First note that P, Q, X and Y are all unit vectors. So, there exist indices i,j such that $p_i \neq 0 \neq q_j$. We see from (3.17)–(3.20) and (3.23)–(3.26) that

$$\begin{aligned} p_i(aC_i + b\phi_i S_i) &= a'p'_i, \\ p_i(a\theta_i S_i - b\theta_i\phi_i C_i) &= b'p'_i, \\ p_i(cC_i + d\phi_i S_i) &= c'p'_i, \\ p_i(c\theta_i S_i - d\theta_i\phi_i C_i) &= d'p'_i, \end{aligned} \tag{3.47}$$

and

$$\begin{aligned} q_j(aC_j + b\phi_j S_j) &= a'q'_j, \\ q_j(a\theta_j S_j - b\theta_j\phi_j C_j) &= b'q'_j, \\ q_j(cC_j + d\phi_j S_j) &= c'\omega' q'_j, \\ q_j(c\theta_j S_j - d\theta_j\phi_j C_j) &= d'\omega' q'_j. \end{aligned} \tag{3.48}$$

Arguing exactly as in the proof of Case (1.2) (of this proposition), we find that $|p'_i| = |p_i|$ and $|q'_j| = |q_j|$. Further, the fact that $S_i, S_j \neq 0$ implies (as before) that $a, b, c, d \neq 0$. Setting $\omega_0 = p_i p_i'^{-1} q_j^{-1} q'_j$, we see that equations (3.47) and (3.48) imply the following identities:

$$\begin{aligned} a(\omega_0 C_i - C_j) + b(\omega_0\phi_i S_i - \phi_j S_j) &= 0, \\ a(\omega_0\theta_i S_i - \theta_j S_j) - b(\omega_0\theta_i\phi_i C_i - \theta_j\phi_j C_j) &= 0, \\ c(\omega_0\omega' C_i - C_j) + d(\omega_0\omega'\phi_i S_i - \phi_j S_j) &= 0, \\ c(\omega_0\omega'\theta_i S_i - \theta_j S_j) - d(\omega_0\omega'\theta_i\phi_i C_i - \theta_j\phi_j C_j) &= 0. \end{aligned}$$

The consistency of the above equations demands that

$$\begin{aligned} &(\omega_0 C_i - C_j)(\omega_0\theta_i\phi_i C_i - \theta_j\phi_j C_j) + (\omega_0\phi_i S_i - \phi_j S_j)(\omega_0\theta_i S_i - \theta_j S_j) = 0, \\ &(\omega_0\omega' C_i - C_j)(\omega_0\omega'\theta_i\phi_i C_i - \theta_j\phi_j C_j) + (\omega_0\omega'\phi_i S_i - \phi_j S_j)(\omega_0\omega'\theta_i S_i - \theta_j S_j) = 0. \end{aligned} \tag{3.49}$$

The fact that $\omega' \neq \pm 1$ enables us to derive the following consequence of the two equations above:

$$\omega_0 = \frac{(\theta_i\phi_i + \theta_j\phi_j)C_i C_j + (\theta_i\phi_j + \theta_j\phi_i)S_i S_j}{\theta_i\phi_i(1 + \omega')}.$$

Substituting this value for ω_0 in eq. (3.49), we get

$$\omega'^2 + 2m_{i,j}\omega' + 1 = 0,$$

where, of course, $m_{i,j}$ is as in the statement of relation (3) in the proposition. It follows that $\mathrm{Re}(\omega') = -m_{i,j}$.

As $X, Y \neq 0$, in a similar way to the previous cases, it follows (from equations (3.27)–(3.34)) that there exist indices i', j' such that $|x_{i'}| = |x_i'| \neq 0 \neq |y_{i'}'| = |y_{i'}|$ and

$$\begin{aligned} ax_{i'} &= (a'C_{i'}' + c'\theta_{i'}'S_{i'}')x_{i'}', \\ bx_{i'} &= (b'C_{i'}' + d'\theta_{i'}'S_{i'}')x_{i'}', \\ cx_{i'} &= (a'\phi_{i'}'S_{i'}' - c'\theta_{i'}'\phi_{i'}'C_{i'}')x_{i'}', \\ dx_{i'} &= (b_{i'}'\phi_{i'}'S_{i'}' - d'\theta_{i'}'\phi_{i'}'C_{i'}')x_{i'}', \end{aligned} \qquad (3.50)$$

and

$$\begin{aligned} ay_{j'} &= (a'C_{j'}' + c'\theta_{j'}'S_{j'}')y_{j'}', \\ b\omega y_{j'} &= (b'C_{j'}' + d'\theta_{j'}'S_{j'}')y_{j'}', \\ cy_{j'} &= (a'\phi_{j'}'S_{j'}' - c'\theta_{j'}'\phi_{j'}'C_{j'}')y_{j'}', \\ d\omega y_{j'} &= (b'\phi_{j'}'S_{j'}' - d'\theta_{j'}'\phi_{j'}'C_{j'}')y_{j'}'. \end{aligned} \qquad (3.51)$$

Again, setting $\omega_0' = x_{i'}x_{i'}'^{-1}y_{j'}^{-1}y_{i'}'$, we find the following consequence of the above sets of equations:

$$\begin{aligned} a'(\omega_0'C_{i'}' - C_{j'}') + c'(\omega_0'\theta_{i'}'S_{i'}' - \theta_{j'}'S_{j'}') = 0, \\ a'(\omega_0'\phi_{i'}'S_{i'}' - \phi_{j'}'S_{j'}') - c'(\omega_0'\theta_{i'}'\phi_{i'}'C_{i'}' - \theta_{j'}'\phi_{j'}'C_{j'}') = 0, \end{aligned} \qquad (3.52)$$

and

$$\begin{aligned} b'(\omega_0'\omega C_{i'}' - C_{j'}') + d'(\omega_0'\omega\theta_{i'}'S_{i'}' - \theta_{j'}'S_{j'}') = 0, \\ b'(\omega_0'\omega\phi_{i'}'S_{i'}' - \phi_{j'}'S_{j'}') - d'(\omega_0'\omega\theta_{i'}'\phi_{i'}'C_{i'}' - \theta_{j'}'\phi_{j'}'C_{j'}') = 0. \end{aligned} \qquad (3.53$$

The consistency of these two sets of equations implies that

$$(\omega_0'C_{i'}' - C_{j'}')(\omega_0'\theta_{i'}'\phi_{i'}'C_{i'}' - \theta_{j'}'\phi_{j'}'C_{j'}') + (\omega_0'\phi_{i'}'S_{i'}' - \phi_{j'}'S_{j'}')(\omega_0'\theta_{i'}'S_{i'}' - \theta_{j'}'S_{j'}') = 0,$$

and

$$(\omega_0'\omega C_{i'}' - C_{j'}')(\omega_0'\omega\theta_{i'}'\phi_{i'}'C_{i'}' - \theta_{j'}'\phi_{j'}'C_{j'}') + (\omega_0'\omega\phi_{i'}'S_{i'}' - \phi_{j'}'S_{j'}')(\omega_0'\omega\theta_{i'}'S_{i'}' - \theta_{j'}'S_{j'}') =$$

and we may deduce as before that $\mathrm{Re}(\omega) = -m_{i',j'}'$; i.e., the relation (3) is satisfiec Finally the proof of (a) is complete.

(b) If condition (0) is satisfied, we may define

$$A = A' = \begin{pmatrix} \theta_1' & 0 \\ 0 & \theta_{\sigma^{-1}(1)} \end{pmatrix}$$

and

$$U = U' = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & P_\sigma \end{pmatrix}$$

and verify that eqs (3.15) to (3.38) are satisfied; and thus, it is indeed true that $(U \otimes A)$ $(\omega, \theta, \phi, C) = W(\omega, \theta', \phi', C')(U' \otimes A')$.

If condition (1) is satisfied, we may define

$$A = \begin{pmatrix} 0 & 1 \\ -\omega\bar{\zeta} & 0 \end{pmatrix}, \quad A' = \begin{pmatrix} 0 & \omega \\ -\omega\bar{\zeta} & 0 \end{pmatrix}$$

and

$$U = \begin{pmatrix} 0 & 1 & 0 \\ \omega & 0 & 0 \\ 0 & 0 & -\theta'\phi' P_\sigma \end{pmatrix}, \quad U' = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & P_\sigma \end{pmatrix},$$

and verify that (3.15) to (3.38) are satisfied; and thus, it is indeed true that $(U \otimes A)W(\omega, \theta, \phi, C) = W(\omega, \theta', \phi', C')(U' \otimes A')$.

(c) (i) By Ocneanu's compactness result (see [O1] or [JS]), we know that

$$A_0^{\infty'} \cap A_1^{\infty} = (M_n \otimes 1) \cap W(M_n \otimes 1)W^*.$$

It is easily seen that if $X = E_{11} \otimes 1$, then $WXW^* = X$, and so we see that $A_0^{\infty'} \cap A_1^{\infty}$ contains a non-trivial projection, thus establishing reducibility of the vertical subfactor.

(ii) In this case, Ocneanu's compactness result says that

$$A_\infty^0 \cap A_\infty^1 = (1 \otimes M_2) \cap W(1 \otimes M_2)W^*.$$

The above algebra does not reduce to the scalars – i.e., the horizontal sub-factor is reducible – precisely when it is possible to find non-scalar matrices

$$X = \begin{pmatrix} x_1 I_n & x_2 I_n \\ x_3 I_n & x_4 I_n \end{pmatrix}, \quad Y = \begin{pmatrix} y_1 I_n & y_2 I_n \\ y_3 I_n & y_4 I_n \end{pmatrix} \in 1 \otimes M_2,$$

where $x_i, y_i \in \mathbb{C}$ such that $WX = YW$.

Easy calculation shows that this matrix equation is satisfied if and only if the following relations hold:

$$\begin{aligned} x_1 &= y_1, x_4 = y_4, \\ x_2 &= y_2 = \omega y_2, x_3 = y_3 = \omega x_3, \\ x_3\theta S &= x_2\phi S, \\ x_2(1+\theta\phi)C &= (x_1 - x_4)\theta S, \\ x_3(1+\theta\phi)C &= (x_1 - x_4)\phi S. \end{aligned} \tag{**}$$

First we will prove that the conditions (1) and (2) are sufficient for the horizontal subfactor to be reducible.

(1) If $S = 0$, it is readily seen that a non-scalar solution to the above system of equations is provided by

$$x_i = y_i = \begin{cases} 1 & \text{if } i = 1 \\ 0 & \text{otherwise.} \end{cases}$$

(2) Suppose $\omega = 1$ and there exists scalar $\lambda_1 \in \mathbb{T}$ and $\lambda_2 \in \mathbb{C}$ such that $\phi S = \lambda_1 \theta S$ and $(1+\theta\phi)C = \lambda_2\theta S$. Choose scalars $x \neq 0$ and x_1, x_4 such that $x_1 - x_4 = \lambda_2 x$. Now if we define

$$X = Y = \begin{pmatrix} x_1 I_n & x I_n \\ \lambda_1 x I_n & x_4 I_n \end{pmatrix},$$

then it is an easy verification to see that the set of equations (**) is satisfied.

Now to prove the necessity of one of the conditions (1) and (2) to hold for the horizontal subfactor to be reducible, we will prove that the horizontal subfactor is irreducible if both the conditions (1) and (2) are not satisfied. So assume $S \neq 0$.

If $\omega \neq 1$, it follows at once from the second and fourth lines of (**) that the equations above are satisfied if and only if $x_2 = y_2 = x_3 = y_3 = 0$, and $x_1 = y_1 = x_4 = y_4$ i.e., if and only if $X = Y = \zeta I_{2n}$ for some $\zeta \in \mathbb{C}$. Hence the horizontal subfactor is irreducible in this case.

Suppose θS and ϕS are not scalar multiples of one another. Then we may deduce from the third line of (**) (as $S \neq 0$) that $x_3 = x_2 = 0$; since $S \neq 0$, either of the last two lines then forces $x_1 = x_4$.

Suppose θS and $(1 + \theta\phi)C$ are not scalar multiples of one another (in particular $(1 + \theta\phi)C \neq 0$). Then we may deduce from the last two lines of (**) (also as $S \neq 0$) that $x_3 = x_2 = 0$ and $x_1 = x_4$. □

We end this section with the following Proposition, which asserts the existence of a continuous $(3n - 6)$-parameter family of pairwise inequivalent connections in $B(2,n)$. It also asserts that the number $(3n - 6)$ is sharp. What we mean by the sharpness of the number $(3n - 6)$ is that there does not exist a subset $\mathcal{B} \subset B(2,n)$ with the following two properties: (i) no two distinct elements of $\mathcal{B}$ are equivalent (as connections); and (ii) $\mathcal{B}$ is homeomorphic to an open subset of Euclidean space of dimension $(3n - 5)$.

PROPOSITION 3.3

There exist non-empty open sets $\Omega \subset \mathbb{T}$, $\Theta \subset \mathbb{T}^{n-2}$, $\Phi_0 \subset \mathbb{T}^{n-3}$, $\Gamma \subset (0,1)^{n-2}$ such that if $(\omega, \theta, \phi, C), (\omega', \theta', \phi' C') \in \Omega \times \Theta \times \Phi \times \Gamma$, where $\Phi = \{1\} \times \Phi_0$ and $(\omega, \theta, \phi, C) \neq (\omega', \theta', \phi' C')$ then $W(\omega, \theta, \phi, C)$ is not equivalent to $W(\omega', \theta', \phi', C')$. Thus, there exist a $(3n - 6)$ parameter family of pairwise inequivalent connections and that is the best possible number.

Further, we may assume that $1 \notin \Omega \cup \Gamma$; hence all these connections have the property that the associated vertical subfactor is reducible and has index n^2, and the horizontal subfactor is irreducible and has index 4.

Proof. Fix $0 < x_1 < x_2 < \pi/4$, and define $\Omega_0 = \{e^{ix} \in \mathbb{T} : x_1 \leq x \leq x_2\}$. Fix $\pi/2 < y_1 < y_2 < 3\pi/4$, such that $0 < y_2 - y_1 < x_1$, and let $\Theta_0 = \{e^{ix} \in \mathbb{T} : y_1 \leq x \leq y_2\}$.

The definitions have the following (easily verified) consequences. Suppose $\omega, \omega' \in \Omega_0$ and $\zeta, \zeta' \in \Theta_0$ are arbitrary. Then,

- $\zeta\zeta' \notin \Omega_0$;
- $\mathrm{Re}(\omega) + \mathrm{Re}(\zeta) \neq 0$;
- $\mathrm{Re}(\omega) + \mathrm{Re}(\bar{\omega}'\zeta) \neq 0$;
- $\mathrm{Re}(\omega) - \mathrm{Re}(\zeta\bar{\zeta}') \neq 0$.

Define $f : [0,1] \times \mathbb{T} \times \mathbb{T} \to \mathbb{R}$ by $f(C, \zeta, \omega) = \mathrm{Re}(\omega) + (1 - C^2) + C^2\,\mathrm{Re}(\zeta)$. Then $0 \notin f(\{1\} \times \Theta_0 \times \Omega_0)$. The compactness of $(\{1\} \times \Theta_0 \times \Omega_0)$ and continuity of f imply the existence of an $\epsilon > 0$ such that for all $C \in (1 - \epsilon, 1), \omega \in \Omega_0, \zeta \in \Theta_0$, we have $\mathrm{Re}(\omega) + (1 - C^2) + C^2\mathrm{Re}(\zeta) \neq 0$. In a similar way, by considering suitable continuous functions, we can see that if ϵ is chosen sufficiently small, then the following relations are also valid.

Suppose $\omega \in \Omega_0$, and $\theta, \phi, \theta', \phi' \in \mathbb{T}$ are such that $\theta\phi, \theta'\phi' \in \Theta_0$, and suppose $C, C' \in (1 - \epsilon, 1)$. Then we simultaneously have

$$\mathrm{Re}(\omega) + (1 - C'^2) + C'^2\,\mathrm{Re}(\bar{\omega}'\theta\phi) \neq 0,$$

and

$$\mathrm{Re}(\omega) + m \neq 0,$$

where $m = 1 - (1 + \mathrm{Re}(\theta\phi\bar{\theta}'\bar{\phi}'))C^2C'^2 - (1 + \mathrm{Re}(\theta\phi'\bar{\theta}'\bar{\phi}))S^2S'^2 - 2(\mathrm{Re}(\theta\bar{\theta}') + \mathrm{Re}(\phi\bar{\phi}'))$ $CC'SS'$.

Let Ω denote the interior of Ω_0, and $\Gamma_0 = (1 - \epsilon, 1)$. Let $\{\Theta_i' : 1 \leq i \leq n-2\}$ be a collection of pairwise disjoint open subsets of Θ_0. Define $\Gamma = \{\mathrm{diag}(C_1, \ldots, C_{n-2}) : C_i \in \Gamma_0 \forall i\}$, and $\Theta' = \{\mathrm{diag}(\zeta_1, \ldots, \zeta_{n-2}) : \zeta_i \in \Theta_i' \forall i\}$.

Define $\Theta_1 = \Theta_1', \Phi_1 = \{1\}$ and for $1 < i \leq n-2$, choose non-empty open subsets $\Theta_i, \Phi_i \subset \mathbb{T}$ such that $\Theta_i\Phi_i \subset \Theta_i'$. Let $\Theta = \{\mathrm{diag}(\theta_1, \ldots, \theta_{n-2}) \in M_{n-2} : \theta_i \in \Theta_i \forall i\}$ and $\Phi = \{\mathrm{diag}(\phi_1, \ldots, \phi_{n-2}) \in M_{n-2} : \phi_i \in \Phi_i \forall i\}$.

Suppose now that $W(\omega, \theta, \phi, C)$ is equivalent to $W(\omega', \theta', \phi', C')$, where $(\omega, \theta, \phi, C)$, $(\omega', \theta', \phi', C') \in \Omega \times \Theta \times \Phi \times \Gamma$.

First notice that if $\zeta, \zeta' \in \Theta'$ and if $\sigma \in S_{n-2}$ are such that $(\mathrm{Ad}(P_\sigma))(\zeta) = \zeta'$, then necessarily $\zeta = \zeta'$ and σ is the identity permutation.

Our choice of (ϵ and consequently of) Γ ensures that neither of the relations (2) or (3) of Proposition 3.2(a) can occur. Suppose the relation (1) were to hold; this would imply that (in the notation of the proposition) $(\mathrm{Ad}(P_\sigma))(\theta\phi) = \omega(\theta'\phi')^*$; in particular, looking at any one diagonal entry of this matrix equation, we would be able to produce elements $\zeta_1, \zeta_2 \in \Theta_0$ such that $\omega = \zeta_1\zeta_2$, which we have already observed to be impossible. Thus the relation (1) can also not hold.

Thus, by Proposition 3.2(a), the relation (0) must necessarily hold. Then the permutation σ (whose existence is the content of (0)) must satisfy the condition $(\mathrm{Ad}(P_\sigma))(\theta\phi) = (\theta'\phi')$, which can only happen when σ is the identity permutation (by the discussion in the paragraph preceding the last one). Hence $\phi = \bar{\zeta}\phi'$, where ζ is as in the statement of Proposition 3.2(a) (0); since $\phi_1 = \phi_1' = 1$, we see that necessarily $\zeta = 1$; but relation (0), when $\sigma = id$ and $\zeta = 1$, then just says that $(\omega, \theta, \phi, C) = (\omega', \theta', \phi', C')$.

Now we will prove that the number $(3n - 6)$ is sharp.

For this, let $F : \mathbb{T} \times \mathbb{T}^{n-2} \times \mathbb{T}^{n-2} \times [0, 1]^{n-2} \longrightarrow B(2, n)$ denote the (obviously continuous) mapping given by $F(\omega, \theta, \phi, C) = W(\omega, \theta, \phi, C)$. Suppose now that there exists a subset $\mathcal{B}$ with the following two properties: (i) no two distinct elements of $\mathcal{B}$ are equivalent (as connections); and (ii) $\mathcal{B}$ is homeomorphic to an open subset of Euclidean space of dimension $(3n - 5)$. Then $F^{-1}(\mathcal{B})$ is a subset of $\mathbb{T}^{2n-3} \times [0, 1]^{n-2}$ which is homeomorphic to an open subset of $\mathbb{T}^{2n-3} \times [0, 1]^{n-2}$, and is consequently itself open (see, for instance [Spa], Th. 4.8.16).

So it suffices to show that any open subset of $\mathbb{T}^{3n-5} \times (0, 1)^{n-2}$ contains two distinct points whose images under F are equivalent, as connections. It clearly suffices to establish this assertion when the open subset is a product $\Omega \times \Theta \times \Phi \times \Gamma$ with open factors.

So, suppose $\Omega \subset \mathbb{T}, \Theta, \Phi \subset \mathbb{T}^{n-2}, \Gamma \subset (0, 1)^{n-2}$ are open subsets, Let $\phi \in \Phi$, $\theta \in \Theta$ be arbitrary. As Φ is assumed to be open, for all $\epsilon > 0$ there exists a $\phi'' \in \Phi$ such that $\phi_1 \neq \phi_1''$ and $\mathrm{Arg}(\phi_1\bar{\phi}_1'') < \epsilon$, where ϕ_1 and $\bar{\phi}_1''$ are the (1,1)th entry of ϕ and ϕ'' respectively. Let $\zeta = \phi_1\phi_1''$. Now define $\theta_i' = \theta_i\zeta$ for $i = 1, 2, \ldots, n-2$ and $\phi_i' = \phi_i\bar{\zeta}$ for $1 = 2, 3, \ldots, n-2$. Now choose ϵ, as Θ and ϕ are open, so that $\theta' \in \Theta$ and $\phi' \in \Phi$. Now it is easily seen that the pair $(\omega, \theta, \phi, C)$ and $(\omega, \theta', \phi', C)$ satisfies the relation (0) in Proposition 3.2, and hence, using (b) of the same proposition, we conclude that $W(\omega, \theta, \phi, C)$ is equivalent to $W(\omega, \theta', \phi', C)$. Finally, the proof is complete. □

4. The principal graph of the horizontal subfactor

In [P] it is shown that for finite-depth subfactors of index 4, the principal graph has to l
one of the extended Dynkin diagrams. We will show that all those diagrams can
obtained from vertex models coming from $B(2,n)$ for some n.

Theorem 4.1 (Popa). *Let $N \subset M$ be an inclusion of II_1 factors, with finite depth a $[M:N]=4$. Then the principal graph for the inclusion $N \subset M$ is one of the followi diagrams: $A_n^{(1)}, D_n^{(1)}, E_6^{(1)}, E_7^{(1)}, E_8^{(1)}$.*

For a group $G \subset U(N)$, let π denotes the standard (or identity) representation of G $U(N)$, and let $C(\hat{G},\pi)$ denote the bipartite graph obtained as follows: let $\mathcal{G}$ denote t bipartite graph with the set of even (respectively odd) vertices being given by $\mathcal{G}^{(0)}$ $\hat{G}\times\{0\}$ (respectively $\mathcal{G}^{(1)}=\hat{G}\times\{1\}$), where $\hat{G}$ denotes the (unitary) dual of G, and t number of bonds joining $(\rho,0)$ and $(\sigma,1)$ is given by $(\rho\otimes\pi,\sigma)$; finally, let $C(\hat{G},$ denote the connected component in $\mathcal{G}$ containing $(\mathrm{tr},0)$, where tr denotes the triv representation of G.

The following theorem is proved in [USC] (also, see [BHJ] and [JS]).

Theorem 4.2. *Let $\{\gamma_1,\gamma_2,\ldots,\gamma_n\}$ be any collection of $k\times k$ unitary matrices, and def $W_{\beta b}^{\alpha a}=\delta_\beta^\alpha(\gamma_\alpha)_b^a$; then W is a biunitary and the principal graph of the horizontal subfac given by the vertex model corresponding to W is $C(G,\pi)$, where G is the group genera by $\{\gamma_1,\gamma_2,\ldots,\gamma_n\}$.*

Suppose H is a finite subgroup of SO(3). Let $\phi:\mathrm{SU}(2)\to\mathrm{SO}(3)$ be the 2-f covering map (i.e., surjective homomorphism such that $\ker\phi=\{+I,-I\}$); let π_n be (unique, up to isomorphism) irreducible representation of SU(2) of dimension $n+1$. $G=\phi^{-1}(H)$, and let $\pi=\pi_1|_G$. The following lemma can be easily seen to be true.

Lemma 4.3. *Let $\rho\in\hat{G}$. Then* (i) $(\rho,0)\in C(\hat{G},\pi)^{(0)}$ *if and only if $\pi(-1)=1$ if and onl $\pi=\pi_0\circ\phi$ for some $\pi_0\in\hat{H}$.* (ii) $(\rho,1)\in C(\hat{G},\pi)^{(1)}$ *if and only if $\pi(-1)\neq 1$ if and o if π does not factor through H.*

PROPOSITION 4.4

For all $\mathcal{G}\in\{A_n^{(1)},D_n^{(1)},E_6^{(1)},E_7^{(1)},E_8^{(1)}\}$ there exists an $n\in\mathbb{N}$ and $W\in B(2,n)$ such that principal graph of the horizontal subfactor given by the vertex model corresponding tc is $\mathcal{G}$.

Proof. It is enough to show that there exists $G\subset SU(2)$ with the property $C(\hat{G},\pi)=\mathcal{G}$. Note that π is self-contragredient and faithful. Using the lemma and sc combinatorial arguments one can see, without too much difficulty, that if we let H be group Z_n, D_n, A_4, S_4 or A_5, then the corresponding Cayley graph $C(\hat{G},\pi)$ turns out tc the extended Coxeter graph $A_{2n}^{(1)}, D_{n+2}^{(1)}, E_6^{(1)}, E_7^{(1)}$, or $E_8^{(1)}$ respectively.

Acknowledgement

I would like to acknowledge my sincere thanks to Prof V S Sunder for suggesting problem, patiently going through the manuscript, pointing out mistakes, and encoura me at every stage. I also would like to thank Dr Vijay Kodiyalam for helping in combinatorial argument used to sort out the graphs in the last section and Dr D S Nag for telling about a Theorem in [Spa] which is used in the proof of Proposition 3.3.

References

[BHJ] Bacher Roland, Harpe de la Pierre and Jones V F R, Carres commutatifs et invariants de structures combinatoires, *Comptes Rendus Acad. Sci. Paris, Serie* 1 (1995) 1049–1054

[GHJ] Goodman F, Harpe de la Pierre and Jones V F R, *Coxeter graphs and towers of algebras* (New York: MSRI Publ., 14, Springer) (1989)

[HS] Haagerup U and Schou J, *Some new subfactors of the hyperfinite II_1 factor*, preprint (1989)

[J] Jones V F R, Index for subfactors, *Invent. Math.* **71** (1983) 1–25

[JS] Jones V and Sunder V S, Introduction to subfactors, *LMS Lect. Note Ser. 123* (Cambridge) (1997)

[O1] Ocneanu A, Quantized groups, String algebras and Galois theory for algebras, *Operator Algebras and Appl. (Warwick) vol. 2 (1987); London Math. Soc. Lecture Notes Ser.* (Cambridge University Press) (1988) 136, pp. 119–172

[P] Popa S, Sur la classification sous-facteurs d'indice fini du facteurs hyperfini, *C R Acad. Sciences*, Serie I Mathematiques 311 **95** (1990) 95–110

[P1] Popa S, Classification of amenable subfactors of type II, *Acta Math.* **172** (1994) 163–255

[Spa] Spanier E H, *Algebraic Topology* (New York: Mc-Graw Hill) (1966)

[Sr] Srnivasan R, *Connections on small vertex models*, thesis, submitted to the Indian Statistical Institute (1998)

[USC] Uma Krishnan, Sunder V S and Varughese Cherian, On some subfactors arising from vertex models, *J. Funct. Anal.* **140** (1996) 449–471

Proc. Indian Acad. Sci. (Math. Sci.), Vol. 110, No. 1, February 2000, pp. 55–59.
© Printed in India

Transformation semigroup compactifications and norm continuity of weakly almost periodic functions

A JALILIAN and M A POURABDOLLAH

Faculty of Mathematical Sciences, Ferdowsi University of Mashhad,
P.O. Box 91775-1159, Mashhad, Iran

MS received 21 June 1999

Abstract. We prove if there exists a separately continuous action of a topologically right simple semitopological semigroup S on a topological space X and if S acts topologically surjective on X then each weakly almost periodic function on X, with respect to S, is left norm continuous.

Keywords. Transformation semigroup; left norm continuous; weakly almost periodic.

Throughout the paper, (S,X) denotes a transformation semigroup, i.e. a semigroup S, a set X, and a map (called the action of S on X) $(s,x) \longrightarrow sx$: $S \times X \longrightarrow X$ such that $s(tx) = (st)x$ for all $s,t \in S$ and $x \in X$. If T is a sub-semigroup of S and Y is a T-invariant subset of X (i.e. $Y \supseteq TY = \{ty : t \in T, y \in Y\}$), then we say (T,Y) is a *sub-transformation semigroup* of (S,X). When S and X are topological spaces, we say (T,Y) is dense in (S,X) if T is dense in S and Y is dense in X.

For notation and terminology we shall follow Berglund *et al* [1], as far as possible. For a topological space $X, \mathcal{C}(X)$ is the C^*-algebra of bounded continuous complex-valued functions on X with supremum norm. For a C^*-subalgebra $\mathcal{F}$ of $\mathcal{C}(X), X^{\mathcal{F}}$ denotes the spectrum of $\mathcal{F}(=$ the set of all multiplicative means on $\mathcal{F})$ which is weak *-compact in the topological dual $\mathcal{F}^*$ of $\mathcal{F}$. The evaluation map $\epsilon : X \longrightarrow X^{\mathcal{F}}$, with a weak *-dense image, is defined by $\epsilon(x)f = f(x)$ $(x \in X, f \in \mathcal{F})$.

For $s,t \in S$ and $x \in X$, we define the translation maps $\lambda_s : S \longrightarrow S, \rho_s : S \longrightarrow S$, $\dot{\lambda}_s : X \longrightarrow X$ and $\dot{\rho}_x : S \longrightarrow X$ by $\lambda_s(t) = st = \rho_t(s)$ and $\dot{\lambda}_s(x) = sx = \dot{\rho}_x(s)$.

When S and X are topological spaces, we say (S,X) is *left topological* if λ_s and $\dot{\lambda}_s$ are continuous for all $s \in S$, *right topological* if ρ_s and $\dot{\rho}_x$ are continuous for all $s \in S$ and $x \in X$, *semitopological* if it is both left and right topological, *topological* if the multiplication in S and the action of S on X are continuous. If there is a separately (resp. jointly) continuous action of a semitopological (resp. topological) group S with identity e on a topological space X and $ex = x$ for all $x \in X$, (S,X) is called a semitopological (resp. topological) transformation group. (S,X) is said to be compact (resp. Hausdorff) if so are S and X.

For $s \in S$ and $x \in X$, we consider the translation operators $L_s = \lambda_s^*$ and $R_s = \rho_s^*$ on $\mathcal{C}(S)$; $\dot{\mathrm{L}}_s = (\dot{\lambda}_s)^*$ and $\dot{\mathrm{R}}_x = (\dot{\rho}_x)^*$ on $\mathcal{C}(X)$, which are bounded with norms ≤ 1. Trivially $\dot{\mathrm{L}}_{st} = \dot{\mathrm{L}}_t\dot{\mathrm{L}}_s, \dot{\mathrm{R}}_{sx} = R_s\dot{\mathrm{R}}_x, L_s\dot{\mathrm{R}}_x = \dot{\mathrm{R}}_x\dot{\mathrm{L}}_s$.

A subset $\mathcal{F}$ of $\mathcal{C}(S)$ (resp. $\mathcal{H}$ of $\mathcal{C}(X)$) is called *translation invariant* if $L_s\mathcal{F} \cup R_s\mathcal{F} \subseteq \mathcal{F}$ (resp. $\dot{\mathrm{L}}_s\mathcal{H} \subseteq \mathcal{H}$) for all $s \in S$.

For a semitopological (S,X), we define

$$\mathcal{AP}(X) = \{f \in \mathcal{C}(X) : \{\dot{\mathrm{L}}_s f : s \in S\} \text{ is norm relatively compact in } \mathcal{C}(X)\}$$
$$\mathcal{WAP}(X) = \{f \in \mathcal{C}(X) : \{\dot{\mathrm{L}}_s f : s \in S\} \text{ is weak relatively compact in } \mathcal{C}(X)\}$$
$$\mathcal{LC}(X) = \{f \in \mathcal{C}(X) : \text{the map } s \rightarrow \dot{\mathrm{L}}_s f : S \rightarrow \mathcal{C}(X) \text{ is norm continuous}\},$$
$$\mathcal{RC}(X) = \{f \in \mathcal{C}(X) : \text{the map } x \rightarrow \dot{\mathrm{R}}_x f : X \rightarrow \mathcal{C}(S) \text{ is norm continuous}\}.$$

All of these function spaces are translation invariant C^*-subalgebras of $\mathcal{C}(X)$ containing the constant functions. Clearly $\mathcal{AP}(X) \subseteq \mathcal{WAP}(X)$.

The following collects some basic facts about these spaces.

PROPOSITION 1.1

Let (S,X) be semitopological. Then

(i) $\mathcal{AP}(X) \subseteq \mathcal{LC}(X) \cap \mathcal{RC}(X)$.
(ii) *If S (resp. X) is compact then $\mathcal{AP}(X) = \mathcal{LC}(X)$ (resp. $\mathcal{RC}(X)$).*
(iii) *If (S, X) is compact then, $\mathcal{AP}(X) = \mathcal{LC}(X) = \mathcal{RC}(X)$ and $\mathcal{WAP}(X) = \mathcal{C}(X)$.*
(iv) *If (S, X) is compact and topological then, $\mathcal{AP}(X) = \mathcal{LC}(X) = \mathcal{RC}(X) = \mathcal{WAP}(X) =$ $\mathcal{C}(X)$.*
(v) *If S (resp. X) is compact and Hausdorff then $\mathcal{C}(X) = \mathcal{RC}(X)$ (resp. $\mathcal{LC}(X)$) if anc only if the action of S on X is (jointly) continuous.*
(vi) *If (S, X) is compact and Hausdorff then $\mathcal{LC}(X) = \mathcal{RC}(X) = \mathcal{C}(X)$ if and only if the action of S on X is continuous.*

Proof. (i) follows from Lemma 1.1(d) of [3]. The proof of (ii) is easy. (iii) follows from (ii) and Lemma 1.1(a) of [3]. (iv) is a consequence of (iii) and Lemma 1.1(b) of [3]. Tc prove (v), note that the action of S on X is continuous if and only if for each $f \in \mathcal{C}(X)$ the map $(s,x) \rightarrow f(sx) : S \times X \rightarrow \mathbb{C}$ is continuous. By ([1], B.3), this is equivalent to norm continuity of $x \rightarrow \dot{\mathrm{R}}_x f : X \rightarrow \mathcal{C}(S)$ (resp. $s \rightarrow \dot{\mathrm{L}}_s f : S \rightarrow \mathcal{C}(X)$) when S (resp. X) is compact. (vi) follows from (v). □

There is no inclusion relation between $\mathcal{WAP}(X)$ and $\mathcal{LC}(X)$ or $\mathcal{RC}(X)$ that holds fo every semitopological (S,X). But if S (resp. X) is compact then, by Proposition 1.1(ii) $\mathcal{LC}(X)$ (resp. $\mathcal{RC}(X)$) $\subseteq \mathcal{WAP}(X)$. Here we are going to obtain some conditions fo establishing the reverse inclusion.

By a *homomorphism* from (S,X) into a transformation semigroup (T,Y) we mean a pair (ϕ,ψ), where $\phi : S \rightarrow T$ is a semigroup homomorphism and $\psi : X \rightarrow Y$ is a map wit the property $\psi(sx) = \phi(s)\psi(x)$ for each $s \in S$ and $x \in X$. We say (ϕ,ψ) is one-to-on (resp. onto) if both ϕ and ψ are one-to-one (resp. onto). A transformation semigrou homomorphism that is one-to-one and onto is called an *isomorphism*. When S and X ar topological spaces, we say (ϕ,ψ) is continuous if both ϕ and ψ are continuous.

By a right (resp. left) topological compactification of a semitopological (S,X) we mea a pair $((\phi,\psi),(T,Y))$, where (T,Y) is a compact Hausdorff right(resp. left) topologica transformation semigroup, and $(\phi,\psi) : (S,X) \rightarrow (T,Y)$ is a continuous homomorphisr such that $(\phi(S),\psi(X))$ is a dense semitopological sub-transformation semigroup c (T,Y). If (T,Y) is semitopological (resp. topological), $((\phi,\psi),(T,Y))$ is a semitopological (resp. topological) compactification of (S,X).

Let $((\phi, \psi), (T, Y))$ and $((\phi', \psi'), (T', Y'))$ be right(resp. left) topological compactifications of (S, X). We call a continuous homomorphism (π, γ) of (T, Y) onto (T', Y') such that $\pi \circ \phi = \phi', \gamma \circ \psi = \psi'$, a *homomorphism* of $((\phi, \psi), (T, Y))$ onto $((\phi', \psi'), (T', Y'))$. If such a homomorphism exists we say $((\phi, \psi), (T, Y))$ is an *extension* of $((\phi', \psi'), (T', Y'))$ and we write

$$(\pi, \gamma) : ((\phi, \psi), (T, Y)) \to ((\phi', \psi'), (T', Y')).$$

If (π, γ) is also one-to-one then it is an *isomorphism* of compactifications.

We call a compactification of (S, X) having a property *P*, a *P-compactification* of (S, X). By a *universal P-compactification* of (S, X) we mean a *P-compactification* of (S, X) that is an extension of every other *P-compactification* of *(S, X)*, which is unique up to isomorphism.

We say a compactification $((\phi, \psi), (T, Y))$ of (S, X) has *the joint continuity property* if the actions of S on T and on Y, i.e. the maps

$$(s, t) \to \phi(s)t : S \times T \to T, (s, y) \to \phi(s)y : S \times Y \to Y,$$

are (jointly) continuous.

For a semitopological (S, X), by an application of Theorem 1.3 of [3], $((\epsilon, \delta), (S^{\mathcal{WAP}}, X^{\mathcal{WAP}})$ (resp. $(S^{\mathcal{AP}}, X^{\mathcal{AP}})))$ is a universal semitopological (resp. topological) compactification of (S, X).

Furthermore, it is shown that $((\epsilon, \delta), (S^{\mathcal{LC}}, X^{\mathcal{LC}})$ (resp. $(S^{\mathcal{RC}}, X^{\mathcal{RC}})))$ is a right(resp. left) topological compactification of (S, X) that is universal with respect to the joint continuity property (see [2]).

Now, we aim to describe our main results. In [4] and [5], some conditions which entail the left norm continuity of weakly almost periodic functions on a transformation semigroup are discussed. Here we are going to give more precise conditions to the same effect. For this we first need a lemma.

Lemma 2.1. *Let (S, X) be a compact Hausdorff semitopological transformation semigroup.*

(i) *If the subsemigroup $T := \{t \in S : tS = S, tX = X\}$ is dense in S, then for each $f \in C(X)$, the map $s \to \dot{L}_s f : S \to C(X)$ is norm continuous at each point of T. Hence the action of S on X is continuous at each point of $T \times X$, and so (T, X) is a topological sub-transformation semigroup of (S, X).*

(ii) *If the subsemigroup $T' := \{t \in S : St = S\}$ is dense in S, then for each $f \in C(X)$, the map $x \to \dot{R}_x f : X \to C(S)$ is norm continuous at each point of $Y := \{y \in X : Sy = X\}$. Hence the action of S on X is continuous at each point of $S \times Y$, and so (T', Y) is a topological sub-transformation semigroup of (S, X).*

Proof. To prove (i), suppose that $f \in C(X), t_0 \in T$ and $\epsilon > 0$. By ([1]; B.1), the function $(s, x) \to f(sx) : S \times X \to \mathbb{C}$ is (jointly) continuous at each point of $\{s_0\} \times X$ for some $s_0 \in S$. Hence, by ([1]; B.3), the set $N := \{s \in S : \|\dot{L}_s f - \dot{L}_{s_0} f\| < \epsilon/2\}$ is a neighborhood of s_0.

By definition of $T, s_0 = t_0 u_0$ for some $u_0 \in S$. Since T is dense in S and $t_0 S = S, t_0 T$ must be dense in S. Choose $t \in T$ such that $t_0 t \in N$. Then $\rho_t^{-1}(N)$ is a neighborhood of t_0, and if $s \in \rho_t^{-1}(N)$ we have

$$
\begin{aligned}
\|\dot{L}_s f - \dot{L}_{t_0} f\| &= \sup_{x\in X} |f(stx) - f(t_0tx)| \\
&\leq \|\dot{L}_{st} f - \dot{L}_{t_0u_0} f\| + \|\dot{L}_{t_0u_0} f - \dot{L}_{t_0t} f\| \\
&< \epsilon/2 + \epsilon/2 = \epsilon,
\end{aligned}
$$

since $s_0 = t_0u_0$ and $st, t_0t \in N$.

Therefore $s \to \dot{L}_s f : S \to \mathcal{C}(X)$ is norm continuous at t_0. By ([1]; B.3), the action of S on X is continuous at each point of $T \times X$.

To prove (ii), let $f \in \mathcal{C}(X), y_0 \in Y$ and $\epsilon > 0$. By ([1]; B.1), the function $(x,s) \to$ $f(sx) : X \times S \to \mathbb{C}$ is (jointly) continuous at each point of $\{x_0\} \times S$ for some $x_0 \in X$ Hence, by ([1]; B.3), the set $B := \{x \in X : \|\dot{R}_x f - \dot{R}_{x_0} f\| < \epsilon/2\}$ is a neighborhood of x_0

The assumptions imply that for any $y_0 \in Y$ there exists an $s_0 \in S$ such that $x_0 = s_0y_0$ and $T'y_0$ is dense in X. Hence $ty_0 \in B$ for some $t \in T'$, and so $\lambda_t^{-1}(B)$ is a neighborhood o y_0. Now if $x \in \lambda_t^{-1}(B)$, then

$$
\begin{aligned}
\|\dot{R}_x f - \dot{R}_{y_0} f\| &= \sup_{s\in S} |f(stx) - f(sty_0)| \\
&\leq \|\dot{R}_{tx} f - \dot{R}_{s_0y_0} f\| + \|\dot{R}_{s_0y_0} f - \dot{R}_{ty_0} f\| \\
&< \epsilon/2 + \epsilon/2 = \epsilon,
\end{aligned}
$$

since $x_0 = s_0y_0$ and $tx, ty_0 \in N$. Thus $x \to \dot{R}_x f : X \to \mathcal{C}(S)$ is norm continuous at y_0. B ([1]; B.3), the action of S on X is continuous at each point of $S \times Y$. □

For a semitopological (S,X), we say S acts *transitively* (resp. *point transitively*) on X i $Sx = X$ for all (resp. for some) $x \in X$. S acts *topologically transitively* on X, if Sx is dens in X for all $x \in X$. Also, we say S acts *surjectively* (resp. *topologically surjectively*) on if $sX = X$ (resp. sX is dense in X) for all $s \in S$.

The following extends Theorem 4.10 of [1] to transformation semigroup setting.

Theorem 2.2. *Let (S,X) be semitopological.*

(i) *If S is topologically right simple and if S acts topologically surjective on X, the $\mathcal{WAP}(X) \subseteq \mathcal{LC}(X)$.*

(ii) *If S is topologically left simple and if S acts topologically transitive on X, the $\mathcal{WAP}(X) \subseteq \mathcal{RC}(X)$.*

Proof. To prove (i), let S be topologically right simple and let S act topologicall surjective on X. Let $((\epsilon, \delta), (T, Y))$ denote the universal semitopological compactificatio of (S,X). Then the subsemigroup $T_1 = \{t \in T : tT = T, tY = Y\}$ of T contains $\epsilon(S)$, an hence by Lemma 2.1(i), for each $f \in \mathcal{C}(Y)$ the map $t \to \dot{L}_t f : T \to \mathcal{C}(Y)$ is nor continuous at each point of $\epsilon(S)$. Since

$$\delta^*(\dot{L}_{\epsilon(s)} f) = \dot{L}_s \delta^*(f)\ (s \in S, f \in \mathcal{C}(Y)),$$

where $\delta^* : \mathcal{C}(Y) \to \mathcal{WAP}(X)$ denotes the dual of the evaluation map $\delta : X \to X^{\mathcal{WAP}}$, follows that

$$\mathcal{WAP}(X) = \delta^*\mathcal{C}(Y) \subseteq \mathcal{LC}(X).$$

The proof of (ii) is similar. □

COROLLARY 2.3

Let (S,X) be a semitopological transformation group. Then $\mathcal{WAP}(X) \subseteq \mathcal{LC}(X)$. In particular if S acts point transitively on X then $\mathcal{WAP}(X) \subseteq \mathcal{LC}(X) \cap \mathcal{RC}(X)$.

Proof. Suppose that (S,X) is a semitopological transformation group. By Theorem 1.1.17 of [1], S is left simple and right simple. Let e be the identity of S and $s \in S, x \in X$, then $x = ex = ss^{-1}x \in sX$ and so $sX = X$. This means that S acts surjectively on X and so, by Theorem 2.2(i), the first assertion holds. To prove the second assertion, let $Sx_0 = X$ for some $x_0 \in X$. Then for each $y \in X$ we have $y = s_0x_0$ for some $s_0 \in S$, and so $Sy = Ss_0x_0 = Sx_0 = X$, i.e. S acts transitively on X. Now the conclusion follows from Theorem 2.2(ii). □

References

[1] Berglund J F, Junghenn H D and Milnes P, *Analysis on Semigroups* (*Function Spaces, Compactifications, Representations*) (New York: Wiley) (1989)

[2] Jalilian A and Pourabdollah M A, Universal compactifications of transformation semigroups, to appear in *J. Sci. I. R. I.*

[3] Junghenn H D, Almost periodic compactifications of transformation semigroups, *Pac. J. Math.* **57(1)** (1975) 207–216

[4] Ebrahimi-Vishki H R and Pourabdollah M A, More on norm-continuity of weakly almost periodic functions, *J. Sci. Univ. Tehran.* **2** (1997) 29–33

[5] Pourabdollah M A, On norm-continuity of weakly almost periodic functions, *J. Sci. Univ. Tehran.* **1** (1996) 39–42

Proc. Indian Acad. Sci. (Math. Sci.), Vol. 110, No. 1, February 2000, pp. 61–68.
© Printed in India

On (N, p, q) summability factors of infinite series

NIRANJAN SINGH and NEETA SHARMA

Department of Mathematics, Kurukshetra University, Kurukshetra 136 119, India

MS received 25 May 1999; revised 18 August 1999

Abstract. In this paper a necessary and sufficient condition has been obtained for $\Sigma\, a_n \epsilon_n$ to be summable $|\bar{N},\ q|$ whenever $\Sigma\, a_n$ is bounded $(N,\, p,\, q)$.

Keywords. Summability factors; conservative matrix.

1. Introduction

Let $\Sigma\, a_n$ be a given infinite series with s_n for its nth partial sum. Let $\{t_n\}$ denote the sequence of (N, p, q) mean of the sequence $\{s_n\}$. The (N, p, q) transform of $s_n = \Sigma_{\nu=0}^{n} a_\nu$ is defined as follows:

$$t_n = \frac{1}{r_n} \sum_{\nu=0}^{n} p_{n-\nu}\, q_\nu\, s_\nu, \tag{1}$$

where

$$r_n = p_0 q_n + \cdots + p_n q_0 \quad (\neq 0)$$
$$p_{-1} = q_{-1} = r_{-1} = 0.$$

Necessary and sufficient conditions for the (N, p, q) method to be regular, that is for $s_n \to s$ to imply $s_n \to s(N, p, q)$ are

(i) $p_{n-\nu} q_\nu / r_n \to 0$ for each integer $\nu \geq 0$ as $n \to \infty$ and
(ii) $\sum_{\nu=0}^{n} |p_{n-\nu} q_\nu| < H |r_n|$, where H is a positive number independent of n.

Let $\{T_n\}$ denote the sequence of $(\bar{N}, q)$ mean of the sequence $\{s_n\}$ defined by

$$T_n = \frac{1}{Q_n} \sum_{\nu=0}^{n} q_\nu s_\nu; \ (Q_n \neq 0) \tag{2}$$

where $Q_n = \sum_{\nu=0}^{n} q_\nu \to \infty$, as $n \to \infty (Q_{-i} = q_{-i} = 0, i \geq 1)$.

We define the sequence of constants $\{c_n\}$ formally by means of the identity

$$\left(\sum_{n=0}^{\infty} p_n x^n \right)^{-1} = \sum_{n=0}^{\infty} c_n x^n, \ c_{-i} = 0, \quad i \geq 1. \tag{3}$$

We also write $c_n^{(1)} = c_0 + c_1 + \cdots + c_n$.

We denote by $\mathcal{M}$, the class of sequences $\{p_n\}$ for which the following holds:

$$p_n > 0, \quad \frac{p_{n+1}}{p_n} \leq \frac{p_{n+2}}{p_{n+1}} \leq 1 \ (n = 0, 1, \ldots). \tag{4}$$

Let $\{p_n\}$ and $\{q_n\}$ be positive sequences. A series $\Sigma\, a_n$ is said to be bounded (N, p, q) or $\Sigma\, a_n = O(1)(N, p, q)$ if

$$\sum_{\nu=1}^{n} p_{n-\nu} q_\nu s_\nu = O(r_n), \quad \text{as } n \to \infty. \tag{5}$$

If X and Y are any two methods of summability, we say (ϵ_n) belongs to the class $[X, Y]$, if $\Sigma\, a_n\epsilon_n$ is summable – Y whenever $\Sigma\, a_n$ is summable X.

Recently Mishra [7], Sarigol and Bor [8] and Sulaiman [9] have obtained summability factor theorems of the type $[|\bar{N}, p_n|_k, |\bar{N}, q_n|_k]$, $[|\bar{N}, p_n|, |\bar{N}, q_n|_k]$, $[|\bar{N}, p_n|_k, |\bar{N}, q_n|]$.

In 1966 Das [2] has proved the following theorem:

Theorem A. *Let $\{p_n\} \in \mathcal{M}$, $q_n \geq 0$. Then if $\Sigma\, a_n$ is summable $|N, p, q|$, it is summable $|\bar{N}, q|$.*

It is therefore, natural to find a summability factor ϵ_n so that $\Sigma\, a_n\epsilon_n$ is summable $|\bar{N}, q|$ whenever $\Sigma\, a_n$ is bounded (N, p, q).

Mazur and Orlicz [5] stated that, if a conservative (i.e. convergence preserving) matrix sums a bounded nonconvergent sequence, then it must sum an unbounded sequence. Zeller [10] obtained a proof of this theorem as a consequence of his study of the summability of slowly oscillating sequences whereas the proof of Mazur and Orlicz [6] was functional analytic, based on rather deep topological properties of FK-spaces. A simple direct proof of this theorem was also given by Fridy [3] which used only the well known Silverman–Toeplitz conditions for regularity.

In view of this remark we state and prove the following summability factor theorem.

Theorem 1. *Let $\{p_n\} \in \mathcal{M}$, $q_0 > 0$, $q_n \geq 0$ and let $\{q_n\}$ be monotonic non-increasing sequence for $n \geq 0$. The necessary and sufficient condition that $\Sigma\, a_n\epsilon_n$ should be summable $|\bar{N}, q|$, whenever*

$$\sum a_n = O(1)(N,p,q), \tag{6}$$

$$\sum_{n=0}^{\infty} \frac{q_n}{Q_n} |\epsilon_n| < \infty, \tag{7}$$

$$\sum_{n=0}^{\infty} |\Delta\epsilon_n| < \infty, \tag{8}$$

$$\sum_{n=0}^{\infty} \frac{Q_{n+1}}{q_{n+1}} |\Delta^2\epsilon_n| < \infty, \tag{9}$$

is that

$$\sum_{n=1}^{\infty} \frac{q_n}{Q_n} |s_n|\, |\epsilon_n| < \infty. \tag{10}$$

Our theorem generalizes and unifies several known results of Mazhar [4], Daniel [1] and others.

2. Lemmas

We need the following lemmas for the proof of our theorem.

Lemma 1 [1]. *Let* $\{q_n\}$ *be positive and monotonic non-increasing sequence. If* $\{\epsilon_n\}$ *is such that*

(i) $\Delta\epsilon_n = o(1)$, *as* $n \to \infty$, *and*

(ii) $\sum \frac{Q_n}{q_{n+1}} |\Delta^2\epsilon_n| < \infty$,

then

$$\sum \frac{Q_n \Delta q_n}{q_n q_{n+1}} |\Delta\epsilon_n| < \infty.$$

Remark. This lemma holds only if $\{q_n\}$ is monotonic non-increasing.

If $\{q_n\}$ is not monotonic non-increasing then conclusion of the lemma may not be true.

Lemma 2 [2]. *Let* $\{p_n\} \in \mathcal{M}$. *Then*

(iii) $c_0 > 0,\ c_n \leq 0\ (n = 1, 2, \ldots)$,

(iv) $\sum_{n=0}^{\infty} c_n x^n$ *is absolutely convergent for* $|x| \leq 1$,

(v) $\sum_{n=0}^{\infty} c_n > 0$,

except when $\sum_{n=0}^{\infty} p_n = \infty$, *in which case*

(vi) $\sum_{n=0}^{\infty} c_n = 0$.

Lemma 3 [2]. *If*

$$t_n = \frac{1}{r_n} \sum_{\nu=0}^{n} p_{n-\nu} q_\nu s_\nu.$$

then

$$s_n = \frac{1}{q_n} \sum_{\nu=0}^{n} c_{n-\nu} r_\nu t_\nu.$$

Lemma 4 [2].

$$\sum_{\mu=0}^{n} c_{n-\mu}^{(1)} r_\mu = Q_n,$$

where c_n, r_n *and* Q_n *are defined as above.*

3. Proof of theorem 1

Let

$$t_n = \frac{1}{r_n} \sum_{\nu=0}^{n} p_{n-\nu} q_\nu s_\nu,$$

and

$$T_n = \frac{1}{Q_n}\sum_{\nu=0}^{n} q_\nu \sum_{r=0}^{\nu} a_r\epsilon_r$$
$$= \frac{1}{Q_n}\sum_{\nu=0}^{n}(Q_n - Q_{\nu-1})a_\nu\epsilon_\nu.$$

Then for $n \geq 1$, we have

$$T_n - T_{n-1} = \frac{q_n}{Q_nQ_{n-1}}\sum_{\nu=0}^{n} Q_{\nu-1}a_\nu\epsilon_\nu.$$

Using Abel's transformation, we get

$$T_n - T_{n-1} = \frac{q_n}{Q_nQ_{n-1}}\left[\sum_{\nu=0}^{n-1} s_\nu\Delta(Q_{\nu-1}\epsilon_\nu) + s_n\epsilon_nQ_{n-1}\right]$$
$$= -\frac{q_n}{Q_nQ_{n-1}}\sum_{\nu=0}^{n-1} q_\nu s_\nu\epsilon_\nu + \frac{q_n}{Q_nQ_{n-1}}\sum_{\nu=0}^{n-1} Q_\nu s_\nu\Delta\epsilon_\nu + \frac{q_n}{Q_n}s_n\epsilon_n.$$

Let

$$T_n - T_{n-1} = \sum_1 + \sum_2 + \sum_3, \quad \text{say.}$$

The theorem is proved if we show that $\Sigma\,|\Sigma_1|$ and $\Sigma\,|\Sigma_2|$ are convergent.

Now

$$\sum_1 = -\frac{q_n}{Q_nQ_{n-1}}\sum_{\nu=0}^{n-1} q_\nu\epsilon_\nu s_\nu$$
$$= -\frac{q_n}{Q_nQ_{n-1}}\sum_{\nu=0}^{n-1} q_\nu\epsilon_\nu \frac{1}{q_\nu}\sum_{\mu=0}^{\nu} c_{\nu-\mu}r_\mu t_\mu$$
$$= -\frac{q_n}{Q_nQ_{n-1}}\sum_{\nu=0}^{n-1} \epsilon_\nu \sum_{\mu=0}^{\nu} c_{\nu-\mu}r_\mu t_\mu$$
$$= -\frac{q_n}{Q_nQ_{n-1}}\sum_{\mu=0}^{n-1} r_\mu t_\mu \sum_{\nu=\mu}^{n-1} c_{\nu-\mu}\epsilon_\nu$$
$$= -\frac{q_n}{Q_nQ_{n-1}}\sum_{\mu=0}^{n-1} r_\mu t_\mu \left[\sum_{\nu=\mu}^{n-1}\left(\sum_{k=0}^{\nu} c_{k-\mu}\right)\Delta\epsilon_\nu + \epsilon_n\sum_{k=0}^{n-1} c_{k-\mu} - \sum_{k=0}^{\mu-1} c\right.$$
$$= -\frac{q_n}{Q_nQ_{n-1}}\sum_{\mu=0}^{n-1} r_\mu t_\mu \left[\sum_{\nu=\mu}^{n-1} c^{(1)}_{\nu-\mu}\Delta\epsilon_\nu + \epsilon_n c^{(1)}_{n-1-\mu}\right]$$
$$= -\frac{q_n}{Q_nQ_{n-1}}\sum_{\mu=0}^{n-1} r_\mu t_\mu \sum_{\nu=\mu}^{n-1} c^{(1)}_{\nu-\mu}\Delta\epsilon_\nu - \frac{q_n}{Q_nQ_{n-1}}\sum_{\mu=0}^{n-1} \epsilon_n c^{(1)}_{n-1-\mu}r_\mu t_\mu$$
$$= \sum_{11} + \sum_{12}, \quad \text{say.}$$

Then as $t_n = O(1)$, using lemmas 2, 3 and 4,

$$\sum_{n=1}^{\infty}\left|\sum_{11}\right| \le \sum_{n=1}^{\infty}\frac{q_n}{Q_nQ_{n-1}}\sum_{\mu=0}^{n-1} r_\mu|t_\mu|\sum_{\nu=\mu}^{n-1} c^{(1)}_{\nu-\mu}|\Delta\epsilon_\nu|, \text{ since } c^{(1)}_{\nu-\mu} > 0 \text{ by Lemma 2.}$$

$$\le K\sum_{n=1}^{\infty}\frac{q_n}{Q_nQ_{n-1}}\sum_{\nu=0}^{n-1}|\Delta\epsilon_\nu|\sum_{\mu=0}^{\nu} c^{(1)}_{\nu-\mu} r_\mu$$

$$= K\sum_{n=1}^{\infty}\frac{q_n}{Q_nQ_{n-1}}\sum_{\nu=0}^{n-1}|\Delta\epsilon_\nu|Q_\nu$$

$$= K\sum_{\nu=0}^{\infty}|\Delta\epsilon_\nu|Q_\nu\sum_{n=\nu+1}^{\infty}\frac{q_n}{Q_nQ_{n-1}}$$

$$= K\sum_{\nu=0}^{\infty}|\Delta\epsilon_\nu| < \infty.$$

Similarly,

$$\sum_{n=1}^{\infty}\left|\sum_{12}\right| \le K\sum_{n=1}^{\infty}\frac{q_n}{Q_nQ_{n-1}}|\epsilon_n|\sum_{\mu=0}^{n-1} c^{(1)}_{n-1-\mu} r_\mu$$

$$= K\sum_{n=1}^{\infty}\frac{q_n}{Q_n}|\epsilon_n| < \infty.$$

Hence

$$\sum_{n=1}^{\infty}\left|\sum_{1}\right| < \infty.$$

Again,

$$\sum_{2} = \frac{q_n}{Q_nQ_{n-1}}\sum_{\nu=0}^{n-1} Q_\nu\Delta\epsilon_\nu\frac{1}{q_\nu}\sum_{\mu=0}^{\nu} c_{\nu-\mu}r_\mu t_\mu$$

$$= \frac{q_n}{Q_nQ_{n-1}}\sum_{\mu=0}^{n-1} r_\mu t_\mu\sum_{\nu=\mu}^{n-1}\frac{Q_\nu}{q_\nu}\Delta\epsilon_\nu c_{\nu-\mu} = \frac{q_n}{Q_nQ_{n-1}}\sum_{\mu=0}^{n-1} r_\mu t_\mu L_1, \quad \text{say},$$

where

$$L_1 = \sum_{\nu=\mu}^{n-1}\frac{Q_\nu}{q_\nu}\Delta\epsilon_\nu c_{\nu-\mu}$$

$$= \sum_{\nu=\mu}^{n-1}\left(\sum_{k=0}^{\nu} c_{k-\mu}\right)\Delta\left(\frac{Q_\nu}{q_\nu}\Delta\epsilon_\nu\right) + \frac{Q_n}{q_n}\Delta\epsilon_n\sum_{k=0}^{n-1} c_{k-\mu}$$

$$= \sum_{\nu=\mu}^{n-1}\left(\sum_{k=\mu}^{\nu} c_{k-\mu}\right)\left[\Delta\left(\frac{Q_\nu}{q_\nu}\right)\Delta\epsilon_\nu + \frac{Q_{\nu+1}}{q_{\nu+1}}\Delta^2\epsilon_\nu\right] + \frac{Q_n}{q_n}\Delta\epsilon_n\sum_{k=\mu}^{n-1} c_{k-\mu}$$

$$= \sum_{\nu=\mu}^{n-1}\left(\sum_{k=\mu}^{\nu} c_{k-\mu}\right)\left[-\frac{q_{\nu+1}}{q_\nu}\Delta\epsilon_\nu + Q_{\nu+1}\Delta\left(\frac{1}{q_\nu}\right)\Delta\epsilon_\nu + \frac{Q_{\nu+1}}{q_{\nu+1}}\Delta^2\epsilon_\nu\right] + \frac{Q_n}{q_n}\Delta\epsilon_n\sum_{k=\mu}^{n-1} c_{k-\mu}$$

$$= \sum_{\nu=\mu}^{n-1}\left[-c^{(1)}_{\nu-\mu}\frac{q_{\nu+1}}{q_\nu}\Delta\epsilon_\nu + c^{(1)}_{\nu-\mu}Q_{\nu+1}\Delta\left(\frac{1}{q_\nu}\right)\Delta\epsilon_\nu + c^{(1)}_{\nu-\mu}\right.$$
$$\left.\times\frac{Q_{\nu+1}}{q_{\nu+1}}\Delta^2\epsilon_\nu\right] + \frac{Q_n}{q_n}\Delta\epsilon_n c^{(1)}_{n-1-\mu}.$$

So

$$\sum_2 = -\frac{q_n}{Q_nQ_{n-1}}\sum_{\mu=0}^{n-1}r_\mu t_\mu\sum_{\nu=\mu}^{n-1}c^{(1)}_{\nu-\mu}\frac{q_{\nu+1}}{q_\nu}\Delta\epsilon_\nu$$
$$+\frac{q_n}{Q_nQ_{n-1}}\sum_{\mu=0}^{n-1}r_\mu t_\mu\sum_{\nu=\mu}^{n-1}c^{(1)}_{\nu-\mu}Q_{\nu+1}\Delta\left(\frac{1}{q_\nu}\right)\Delta\epsilon_\nu$$
$$+\frac{q_n}{Q_nQ_{n-1}}\sum_{\mu=0}^{n-1}r_\mu t_\mu\sum_{\nu=\mu}^{n-1}c^{(1)}_{\nu-\mu}\frac{Q_{\nu+1}}{q_{\nu+1}}\Delta^2\epsilon_\nu + \frac{q_n}{Q_nQ_{n-1}}\sum_{\mu=0}^{n-1}r_\mu t_\mu c^{(1)}_{n-1-\mu}\frac{Q_n}{q_n}\Delta\epsilon_n$$
$$=\sum_{21}+\sum_{22}+\sum_{23}+\sum_{24},\quad \text{say.}$$

Therefore to show that

$$\sum_{n=1}^{\infty}\left|\sum_2\right| < \infty$$

it is enough to show that

$$\sum_{n=1}^{\infty}\left|\sum_{2i}\right| < \infty,\quad i = 1,2,3,4.$$

Now as $c^{(1)}_k > 0$ for $k > 0$ and as $t_n = O(1)$

$$\sum_{n=1}^{\infty}\left|\sum_{21}\right| \le \sum_{n=1}^{\infty}\frac{q_n}{Q_nQ_{n-1}}\sum_{\mu=0}^{n-1}r_\mu|t_\mu|\sum_{\nu=\mu}^{n-1}c^{(1)}_{\nu-\mu}\frac{q_{\nu+1}}{q_\nu}|\Delta\epsilon_\nu|$$
$$\le K\left[\sum_{n=1}^{\infty}\frac{q_n}{Q_nQ_{n-1}}\sum_{\mu=0}^{n-1}r_\mu\sum_{\nu=\mu}^{n-1}c^{(1)}_{\nu-\mu}\frac{q_{\nu+1}}{q_\nu}|\Delta\epsilon_\nu|\right]$$
$$= K\left[\sum_{n=1}^{\infty}\frac{q_n}{Q_nQ_{n-1}}\sum_{\nu=0}^{n-1}\frac{q_{\nu+1}}{q_\nu}|\Delta\epsilon_\nu|\sum_{\mu=0}^{\nu}r_\mu c^{(1)}_{\nu-\mu}\right]$$
$$= K\sum_{n=1}^{\infty}\frac{q_n}{Q_nQ_{n-1}}\sum_{\nu=0}^{n-1}\frac{q_{\nu+1}}{q_\nu}|\Delta\epsilon_\nu|Q_\nu$$
$$= K\sum_{\nu=0}^{\infty}\frac{q_{\nu+1}}{q_\nu}Q_\nu|\Delta\epsilon_\nu|\sum_{n=\nu+1}^{\infty}\frac{q_n}{Q_nQ_{n-1}}$$
$$= K\sum_{\nu=0}^{\infty}\frac{q_{\nu+1}}{q_\nu}Q_\nu|\Delta\epsilon_\nu|\frac{1}{Q_\nu} = K\sum_{\nu=0}^{\infty}\frac{q_{\nu+1}}{q_\nu}|\Delta\epsilon_\nu|$$
$$\le K\sum_{\nu=0}^{\infty}|\Delta\epsilon_\nu| < \infty,\quad \text{as } \{q_n\} \text{ is non-increasing.}$$

Note that K is a positive constant not necessarily same at each occurance.

Similarly

$$\begin{aligned}
\sum_{n=1}^{\infty}\left|\sum_{22}\right| &\le \sum_{n=1}^{\infty}\frac{q_n}{Q_nQ_{n-1}}\sum_{\mu=0}^{n-1} r_\mu|t_\mu|\sum_{\nu=\mu}^{n-1} c^{(1)}_{\nu-\mu}\frac{Q_{\nu+1}\Delta q_\nu}{q_\nu q_{\nu+1}}|\Delta\epsilon_\nu| \\
&\le K\sum_{n=1}^{\infty}\frac{q_n}{Q_nQ_{n-1}}\sum_{\mu=0}^{n-1} r_\mu\sum_{\nu=\mu}^{n-1} c^{(1)}_{\nu-\mu}\frac{Q_{\nu+1}\Delta q_\nu}{q_\nu q_{\nu+1}}|\Delta\epsilon_\nu| \\
&= K\sum_{n=1}^{\infty}\frac{q_n}{Q_nQ_{n-1}}\sum_{\nu=0}^{n-1}\frac{Q_{\nu+1}\Delta q_\nu}{q_\nu q_{\nu+1}}|\Delta\epsilon_\nu|\sum_{\mu=0}^{\nu} r_\mu c^{(1)}_{\nu-\mu} \\
&= K\sum_{n=1}^{\infty}\frac{q_n}{Q_nQ_{n-1}}\sum_{\nu=0}^{n-1}\frac{Q_{\nu+1}Q_\nu\Delta q_\nu}{q_\nu q_{\nu+1}}|\Delta\epsilon_\nu| \\
&= K\sum_{\nu=0}^{\infty}\frac{Q_{\nu+1}Q_\nu\Delta q_\nu}{q_\nu q_{\nu+1}}|\Delta\epsilon_\nu|\sum_{n=\nu+1}^{\infty}\frac{q_n}{Q_nQ_{n-1}} \\
&= K\sum_{\nu=0}^{\infty}\frac{Q_{\nu+1}\Delta q_\nu}{q_\nu q_{\nu+1}}|\Delta\epsilon_\nu| < \infty, \text{ by Lemma 1,}
\end{aligned}$$

$$\begin{aligned}
\sum_{n=1}^{\infty}\left|\sum_{23}\right| &\le K\sum_{n=1}^{\infty}\frac{q_n}{Q_nQ_{n-1}}\sum_{\mu=0}^{n-1} r_\mu\sum_{\nu=\mu}^{n-1} c^{(1)}_{\nu-\mu}\frac{Q_{\nu+1}}{q_{\nu+1}}|\Delta^2\epsilon_\nu| \\
&= K\sum_{n=1}^{\infty}\frac{q_n}{Q_nQ_{n-1}}\sum_{\nu=0}^{n-1}\frac{Q_{\nu+1}}{q_{\nu+1}}|\Delta^2\epsilon_\nu|\sum_{\mu=0}^{\nu} r_\mu c^{(1)}_{\nu-\mu} \\
&= K\sum_{\nu=0}^{\infty}\frac{Q_{\nu+1}Q_\nu}{q_{\nu+1}}|\Delta^2\epsilon_\nu| \times \frac{1}{Q_\nu} \\
&= K\sum_{\nu=0}^{\infty}\frac{Q_{\nu+1}}{q_{\nu+1}}|\Delta^2\epsilon_\nu| < \infty
\end{aligned}$$

and

$$\begin{aligned}
\sum_{n=1}^{\infty}\left|\sum_{24}\right| &\le K\sum_{n=1}^{\infty}\frac{q_n}{Q_nQ_{n-1}}|\Delta\epsilon_n|\frac{Q_n}{q_n}\sum_{\mu=0}^{n-1} r_\mu c^{(1)}_{n-1-\mu} \\
&= K\sum_{n=1}^{\infty}\frac{1}{Q_{n-1}}|\Delta\epsilon_n|Q_{n-1} \\
&= K\sum_{n=1}^{\infty}|\Delta\epsilon_n| < \infty.
\end{aligned}$$

Hence

$$\sum_{n=1}^{\infty}\left|\sum_{2}\right| < \infty$$

and the proof of the theorem is completed.

Theorem 2. *Let $\{p_n\} \in \mathcal{M}, q_0 > 0, q_n \ge 0$ and let $\{q_n\}$ be monotonic non-increasing sequence for $n \ge 0$. The necessary and sufficient condition that $\Sigma\, a_n\epsilon_n$ should be* *summable* $|\bar{N},q|$ *whenever*

$$\sum a_n = O(\mu_n)(N,p,q),$$

where $\{\mu_n\}$ is positive and monotonic non-decreasing and $\{\epsilon_n\}$ is such that

$$\sum_{n=0}^{\infty} \frac{q_n}{Q_n} |\epsilon_n| \mu_n < \infty,$$

$$\sum_{n=0}^{\infty} |\Delta\epsilon_n| \mu_n < \infty,$$

$$\sum_{n=0}^{\infty} \frac{Q_{n+1}}{q_{n+1}} |\Delta^2\epsilon_n| \mu_n < \infty,$$

is that

$$\sum_{n=1}^{\infty} \frac{q_n}{Q_n} |s_n|\, |\epsilon_n| < \infty.$$

The proof of theorem 2 is similar to that of theorem 1.

References

[1] Daniel E C, On the $|\bar{N},p_n|$ summability of infinite series, *J. Math. Univ. Jabalpur* **2** (1966) 48

[2] Das G, On some methods of summability, *Q. J. Math. Oxford* **17** (1966) 244–256

[3] Fridy J A, A simple proof of the Mazur–Orlicz summability theorem, *Math. Proc. Ca* *Philos. Soc.* **89** (1981) 391–392

[4] Mazhar S M, $|\bar{N},p_n|$ summability factors of infinite series, *Kodai Math. Sam. Rep.* **18** (19 96–100

[5] Mazur S and Orlicz W, Sur les methodes lineares de sommation, *C.R. Acad Sci. Paris* (1933) 32–34

[6] Mazur S and Orlicz W, On linear methods of summability, *Studia Math.* **14** (1954) 129–

[7] Mishra K N, Absolute summability factors of type (X,Y), *Proc. Am. Math. Soc.* **122** (19 531–539

[8] Sarigol M A and Bor H, Characterization of absolute summability factors, *J. Math. A* *Appl.* **195** (1995) 537–545

[9] Sulaiman W T, Relations on some summability methods, *Proc. Am. Math. Soc.* **118** (19 1139–1145

[10] Zeller K, Factorfolgen bei Limitierungsverfahren, *Math. Z.* **56** (1952) 134–151

Proc. Indian Acad. Sci. (Math. Sci.), Vol. 110, No. 1, February 2000, pp. 69–77.
© Printed in India

Construction of 'Wachspress type' rational basis functions over rectangles

P L POWAR and S S RANA

Department of Mathematics and Computer Science, R.D. University, Jabalpur 482001, India

MS received 15 April 1998; revised 8 October 1999

Abstract. In the present paper, we have constructed rational basis functions of C^0 class over rectangular elements with wider choice of denominator function. This construction yields additional number of interior nodes. Hence, extra nodal points and the flexibility of denominator function suggest better approximation.

Keywords. Rational basis functions; C^0 approximation.

1. Introduction

The welknown wedge construction over rectangular elements of C^0 class in R^2 has been discussed by Ciarlet [1] (see also [4]) where $(n+1)^2$ monomials $x^i y^j, i, j \leq n$ are used as basis functions. Rational basis functions over convex quadrilaterals that are not rectangles were introduced by Wachspress [8]. This concept was used in [5] to achieve C^1-approximation of degree two and three over the triangular mesh of planar region (see also [7]).

Let P_n be the space of polynomials of degree n and Q_n be the space generated by the monomials then $P_n \subset Q_n \subset P_{2n}$ (see [1], p. 56). Here we construct rational basis functions of C^0 class over the same element which are ratios of polynomials of degree $2n$ and $n-1$, respectively. If Q_n^* denotes the space spanned by the rational basis functions then, $P_n \subset Q_n^* \subset P_{n+2}$ which is precisely a restricted class than that of P_{2n} for $n \geq 3$. The extra flexibility of the denominator polynomials and the additional number of interior nodes suggest favourable approximation properties of these functions (see [6]).

We first demonstrate the wedge construction and then show that our rational basis functions satisfy all the properties of wedges specified in [8]. In particular, we have also illustrated an example for $n = 3$ which gives the clear idea of our construction.

2. Notations and some preliminary results

The vertices a_i of a closed convex rectangle K in R^2 are labeled so that a_i and a_{i+1} are consecutive for $i = 1, 2, 3, 4$. For each subset $\mathcal{A}$ of $R^2, P_n^*(\mathcal{A})$ is the $\mathcal{A}$-restriction of the vector space of bivariate polynomial functions of degree $\leq n$ in each of the two variables. Let d_i be the straight line passing through the points a_i and a_{i+1} given by the equation $l_i(x, y) = l_i(\text{say}) = 0, i = 1, 2, 3, 4$. Let $a_{ij}, j = 1, 2 \ldots, n-1$ be any distinct points on the edge $[a_i, a_{i+1}]$. Suppose A and B are points on the lines d_1 and d_4 that do not coincide with a_1. Furthermore, suppose that the line through A and B given by $l(x, y) = l(\text{say}) = 0$ intersects d_3 and d_2 at points C and D, respectively (cf. figure 1). For arbitrary but fixed

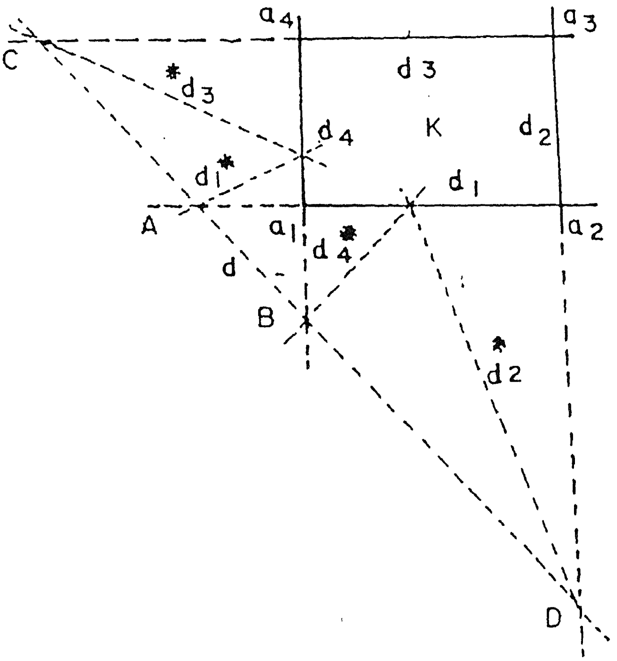


Figure 1. Rectangular element with rational basis.

$j = 1,2,\ldots,n-1$, let d_1^*, d_2^*, d_3^* and d_4^* denote the straight lines passing through respective pairs of points $(a_{4j},A),(a_{1j},D),(a_{4j},C)$ and (a_{1j},B), and let $l_i^*(x,y$ $l_i^*(\text{say}) = 0$ be their corresponding defining equations. Let $\alpha_{ij}(x,y) = \alpha_{ij}(\text{say})$ irreducible conics passing through the points a_{i-1j}, a_{ij}, p and q. The points p an depend on i. For $i = 1,2; i = 2,3;$ $i = 3,4;$ and $i = 4,1$ we choose $p = A$; $q = D$; $p =$ and $q = B$, respectively (see figure 1). We use the usual convention regarding subscripts and throughout this paper, we shall use the notations and definitions give [8] unless stated otherwise.

Remark 2.1. It is essential to construct α_{ij} irreducible. If it degenerates into the pro of two linear forms the denominator function may become the factor of numerator and basis functions will not be of the desired class. Since points A, B, C, D and the edge n are arbitrary, it is always possible to choose four points such that no three of them collinear (see [2] and also [3]) in order to get α_{ij} irreducible.

For the demonstration of the wedge properties of the basis functions which woul defined in the next section, we need the following results of [8].

Lemma 2.1. *If three lines intersect at a point then the ratio of the linear forms w vanish on any two of these lines is constant on the third line.*

Lemma 2.2. *Let P_n, Q_m and L_1 have s distinct triple points then*

$$\frac{P_n(x,y)}{Q_m(x,y)} \equiv \frac{P^1_{n-s}(v)}{Q^1_{m-s}(v)} \operatorname{mod} L_1,$$

where polynomials P^1 and Q^1 are derived from P_n and Q_m by elimination of x or y on line L_1.

Lemma 2.3. *Let Q be a polynomial in x and y which is a product of distinct irreducible factors and let P be a polynomial which is not identically zero. If $P \equiv 0 \bmod Q$, then $Q(x, y)$ must be a factor of $P(x, y)$.*

3. Construction of wedges and their properties

With the notations described in § 2, we define functions $W_i(x,y)$ and $W_{ij}(x,y)$ for $i = 1,2,3,4$ and $j = 1,2,\ldots,n-1$ for the vertices a_i and edge nodes a_{ij} respectively. For each $(x,y) \in K$ and $n \geq 1$

$$W_i(x,y) = \frac{H_i l_{i+1} l_{i+2} \prod_{j=1}^{n-1} \alpha_{ij}}{l^{n-1}}, \tag{3.1}$$

$$W_{ij}(x,y) = \frac{H_{ij} l_i^* l_{i+1} l_{i+2} l_{i+3} \prod_{k=1\, k\neq j}^{n-1} \alpha_{ik}}{l^{n-1}}, \tag{3.2}$$

where H_i and H_{ij} are suitable normalizing constants to ensure that $W_i(a_i) = 1$, $W_{ij}(a_{ij}) = 1$, $i = 1,2,3,4; j = 1,2,\ldots,n-1$.

With the applications of Lemmas 2.1–2.3, we now show that properties described in article 1.5 of [8] are satisfied by the functions (3.1) and (3.2). It is clear from the construction that $W_i(x,y)$ vanishes on the edges opposite of the vertex a_i and at all the nodes a_{ij} and at vertices a_k $(k = 1,2,3,4)$ $k \neq i$. Similarly $W_{ij}(x,y)$ vanishes on the edges opposite of the nodes a_{ij} and at all the vertices a_i and at nodes a_{is} $(s = 1,2,\ldots,n-1)$ $s \neq j$. Another important property may be formulated as follows.

Theorem 3.1. *The restriction of the wedges $W_i(x,y)$ and $W_{ij}(x,y)$ to their respective adjacent edges are polynomials of degree n.*

Proof. Edges d_{i-1} and d_i are adjacent to the vertex a_i and the edge d_i is adjacent to the edge nodes a_{ij}. Consider $R_s \in P_s^*(R^2)$ defined in § 2.

For convenience, we introduce the following notations,

$$\left.\begin{array}{c} \prod_{j=1}^{n-1} \alpha_{ij}(x,y) = R_{2n-2}, \prod_{k=1\, k\neq j}^{n-1} \alpha_{ik}(x,y) = R_{2n-4} \\ l^{n-1}(x,y) = R_{n-1}, l^{n-2}(x,y) = R_{n-2} \end{array}\right\}. \tag{3.3}$$

The algebraic curves R_{2n-2}, R_{n-1} defined in (3.3) and a straight line d_i intersect at a point and a point of intersection is a multiple point of R_{2n-2} and R_{n-1} with multiplicity $(n-1)$. In view of Lemma 2.2, we have

$$\frac{R_{2n-2}}{R_{n-1}} \equiv R_{n-1} \bmod d_i. \tag{3.4}$$

Since K is rectangle,

$$l_{i+2} \equiv C_1 \bmod d_i, \tag{3.5}$$

where C_1 is some constant. Relation (3.4) together with (3.5) gives the following:

$$W_i \equiv R_{n-1} l_{i+1} \bmod d_i$$

or

$$W_i \equiv R_n \bmod d_i.$$

By the similar argument, we may show that

$$W_i \equiv R_n \bmod d_{i-1}.$$

For the other case, the algebraic curves R_{2n-4}, R_{n-2} given by (3.3) and a straight line d_i intersect at a point and a point of intersection is a multiple point of R_{2n-4}, R_{n-2} with multiplicity $(n-2)$. Again we get

$$\frac{R_{2n-4}}{R_{n-2}} \equiv R_{n-2} \bmod d_i \tag{3.6}$$

when we appeal to lemma 2.2. The three straight lines d_i^*, d and d_i intersect at a point, in view of lemma 2.1, we have

$$\frac{l_i^*}{l} \equiv C_2 \bmod d_i, \tag{3.7}$$

$$l_{i+2} \equiv C_3 \bmod d_i (: K \text{ is rectangle}), \tag{3.8}$$

where C_2 and C_3 are some constants. The following congruence relation follows from (3.6), (3.7) and (3.8).

$$W_{ij} \equiv R_{n-2} l_{i+1} l_{i+3} \bmod d_i$$

or

$$W_{ij} \equiv R_n \bmod d_i.$$

This completes the proof of Theorem 3.1.

DEFINITION 3.1

Let $Q_n^*(K)$ be the vector space generated by the function $W_i\,(i = 1, 2, 3, 4)$ and $W_{ij}\,(i = 1, 2, 3, 4;\ j = 1, 2, \ldots, n-1)$ defined by (3.1) and (3.2).

DEFINITION 3.2

Let Σ_n be the set of linearly independent linear forms defined over the space Q_n^* and is given by

$$\sum_n = (v \to v(a_i);\ v \to v(a_{ij}) : i = 1, 2, 3, 4;\ j = 1, 2, \ldots, n-1). \tag{3.9}$$

Denoting a finite element of our construction by (K, Q_n^*, Σ_n), we are now set to prove our next results.

Theorem 3.2. *The finite element* (K, Q_n^*, Σ_n) *given by definitions* 3.1 *and* 3.2 *is of* C^0*-class.*

Proof. Let D^* be an open bounded subset polygon in R^2 and let D_h^* be a triangulation of D^* by rectangles K defined in § 2. Let V_h be a finite element space whose generic element

(K, Q_n^*, Σ_n) is given by (3.1), (3.2) and (3.9). Considering two adjacent rectangles K_i and K_j with common side $K' = [a_i, a_{i+1}]$ and $v \in V_h$, we have the following,

$$v|_{K_i} = p_i \in Q^*(K_i); \quad v|_{K_j} = p_j \in Q^*(K_j).$$

It follows from Theorem 3.1 that the restriction to K' of the basis functions are elements of $P_n^*(K')$ and we get

$$(p_i - p_j)|_{K'} \in P_n^*(K'). \tag{3.10}$$

Since $(n+1)$ data points are prescribed on the rectangular edges, in view of (3.10), we have

$$(p_i - p_j)|_{K'} = 0.$$

This is the desired result.

Theorem 3.3. *Assuming respectively the definitions* 3.1 *and* 3.2 *of* Q_n^* *and* Σ_n, (K, Q_n^*, Σ_n) *is a finite element.*

Proof. Excluding the interior nodes, it may be observed that

$$\dim Q_n^* = \dim \Sigma_n = 4n.$$

It is sufficient, if we prove that the set Σ_n is Q_n^*-unisolvent. In fact, we show that the functions W_i and W_{ij} defined by (3.1) and (3.2) are basis functions of Q_n^* with respect to Σ_n. For i, $k = 1, 2, 3, 4$,

$$W_i(a_k) = \delta_{ik} = [1 \text{ if } i = k;\ 0 \text{ otherwise}]. \tag{3.11}$$

For $i, r = 1, 2, 3, 4$; $j, s = 1, 2, \ldots, n-1$, we have

$$W_{ij}(a_{rs}) = \delta_{ir}\delta_{js} = [1 \text{ if } i = r,\ j = s;\ 0 \text{ otherwise}], \tag{3.12}$$

$$W_i(a_{rs}) = W_{ij}(a_k) = 0. \tag{3.13}$$

The linear independence of the functions defined in (3.1) and (3.2) follows by the relations (3.11), (3.12) and (3.13). Hence, these are the basis functions of the space Q_n^* and the proof is completed.

Remark 3.1. Since our basis functions consists of polynomials of degrees $(2n, n-1)$, we may introduce $(2n-2)(2n-3)/2$ interior nodes to get degree n-approximation. The basis function at each interior node can be defined as the product of the four boundary linear functions with the algebraic curve of degree $2n-4$ which passes through $(2n-2)$ $(2n-3)/2 - 1$ interior nodes in the numerator and the denominator given in that of (3.1) and (3.2). In fact, it may be observed that the set S^* formed by the wedges at the interior nodes together with the wedges defined over the boundary of the rectangle is linearly independent and thus forms the basis for Q_n^* and the theorem 3.2 holds with respect to S^* also.

DEFINITION 3.3

A rational approximation is said to have degree n if every polynomial of greatest degree n can be written as linear combination of the functions (3.1) and (3.2) (see [4], p. 83).

Theorem 3.4. *The finite element given by the definitions* 3.1 *and* 3.2 *is of degree n.*

Proof. Let $\Phi_n(x,y)$ be an element of $P_n^*(K)$ and consider the function $V_n(x,y)$ defi over K by

$$V_n(x,y) = \Phi_n(x,y) - [\Sigma_i \, \Phi_n(a_i)W_i(x,y) + \Sigma_{i,j}\, \Phi_n(a_{ij})W_{ij}(x,y)]. \qquad (3.$$

The function $V_n(x,y)$ must be of the form

$$V_n(x,y) = \frac{R_{2n}}{R_{n-1}}$$

and

$$R_{2n} = B_4 \beta_{2n-4},$$

where B_4 is the boundary curve which vanishes on the rectangle boundary and in v of remark 3.1, the algebraic curve β_{2n-4} is identically zero at all the interior noc Therefore, R_{2n} is the zero polynomial when we appeal to Lemma 2.3. We thus h $V_n(x,y) = 0$. It follows from (3.14) that $\Phi_n(x,y) \in Q_n^*$ and the proof is completed.

4. Importance and relevance of the new construction

(a) We have already mentioned in our remark 2.1 that the conic α_{ij} should be irreduci α_{ij} may also be expressed as the product of two straight lines in which the denominato the wedge functions is one of the factors. With this choice of α_{ij}, our construction redu to the standard basis which span monomials (tensor product polynomials). Hei monomial functions are special case of our rational basis functions.

(b) In case of monomials the number of interior nodes is $(n-1)^2$ to achieve degre approximation whereas in our construction, we get $(2n-2)(2n-3)/2$ interior node achieve the same degree of approximation. Therefore, better approximations are expec with these additional $(n-1)(n-2), (n \geq 3)$ interior nodes.

5. Illustration

In this section, we discuss the following example for special choice of n.

Example 5.1. Consider $n = 3$. Refer figure 2.

Let K be the square with the vertices $a_1(0,0), a_2(1,0), a_3(1,1), a_4(0,1)$ and side nc $a_{11}(1/3,0), a_{12}(2/3,0), a_{21}(1,1/3), a_{22}(1,2/3), a_{31}(2/3,1), a_{32}(1/3,1), a_{41}(0,2/3)$ $a_{42}(0,1/3)$. Let $b_m(m = 1,2,\ldots,6)$ are distinct interior nodes (no three of them collinear) situated at the interior of K and are specified by their respective carte coordinates, $b_1(1/3,1/3)$, $b_2(1/2,1/3)$, $b_3(2/3,1/2)$, $b_4(1/2,1/2)$, $b_5(1/3,2/3)$, $b_6(1/9,2/3)$. The linear forms of d_i and d_i^* for $i = 1,2,3,4$ are given by

$$l_1 \cong y = 0; \quad l_2 \cong x - 1 = 0; \quad l_3 \cong y - 1 = 0,$$

$$l_4 \cong x = 0; \quad l_1^* \cong 3y - x - 1 = 0; \quad l_2^* \cong y + 3x - 1 = 0,$$

$$l_3^* \cong 3y + x - 1 = 0; \quad l_4^* \cong y - 3x + 1 = 0. \qquad ($$

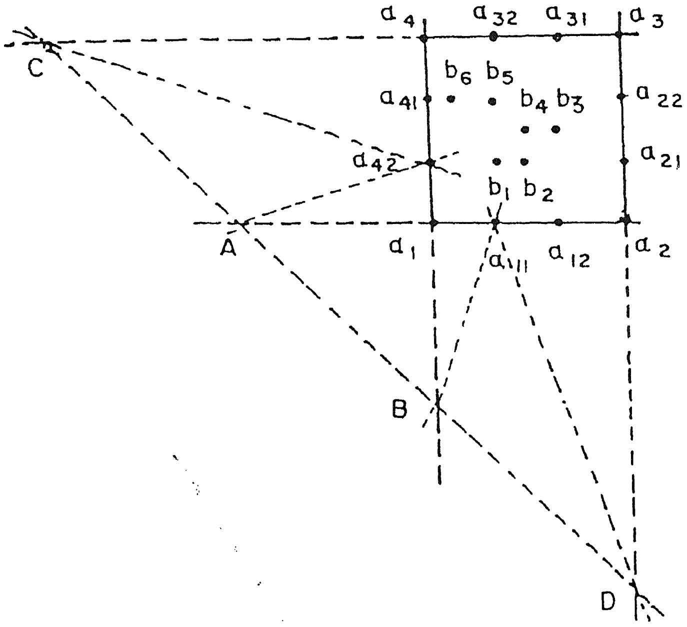


Figure 2. Case when $n = 3$.

If we choose $A(-1, 0)$, $B(0, -1)$, the denominator function $l(x, y)$ reduces to

$$l(x, y) \cong x + y + 1 = 0. \tag{5.2}$$

Let $W^*_m(b_r)$ be the basis elements defined at the interior nodes b_r for $r = 1, 2, \ldots, 6$ such that

$$W^*_m(b_r) = \delta_{mr}; \quad m, r = 1, 2, \ldots, 6,$$

where δ_{mr} is usual Kronecker's delta. We determine conics α_{ij} which are defined in § 2. Suppose

$$\alpha_{ij}(x, y) \cong a_0x^2 + a_1y^2 + a_3x + a_4y + 1 = 0. \tag{5.3}$$

We get the following conics in view of (5.3)

$$\begin{aligned}
\alpha_{11}(x, y) &\cong 6x^2 + 3y^2 + 4x + y - 2 = 0,\\
\alpha_{12}(x, y) &\cong 3x^2 + 6y^2 + x + 4y - 2 = 0,\\
\alpha_{21}(x, y) &\cong 3x^2 - 6y^2 + 4x - 2y + 4 = 0,\\
\alpha_{22}(x, y) &\cong 6x^2 - 3y^2 + 2x - 4y - 4 = 0,\\
\alpha_{31}(x, y) &\cong 6x^2 + 3y^2 + 8x + 5y - 16 = 0,\\
\alpha_{32}(x, y) &\cong 3x^2 + 6y^2 + 5x + 8y - 16 = 0,\\
\alpha_{41}(x, y) &\cong 3x^2 - 6y^2 + 2x - 10y - 1 = 0,\\
\alpha_{42}(x, y) &\cong 6x^2 - 3y^2 + 10x - 2y + 1 = 0.
\end{aligned} \tag{5.4}$$

We now suppose that β_m be the irreducible conic passing through the points b_r (n of b_r's are collinear) such that $m \neq r$ for $m, r = 1, 2, \ldots, 6$. Let

$$\beta_m(x, y) \cong a_0x^2 + a_1y^2 + a_2xy + a_3x + a_4y + 1 = 0.$$

In view of (5.5), we obtain the following conics:

$$\begin{aligned}
\beta_1(x, y) &\cong 36x^2 + 42y^2 - 120x - 113y + 156xy + 58 = 0,\\
\beta_2(x, y) &\cong 9x^2 - 18y^2 - 30x + 5y + 39xy + 5 = 0,\\
\beta_3(x, y) &\cong 18x^2 + 21y^2 - 22x - 28y + 21xy + 10 = 0,\\
\beta_4(x, y) &\cong 54x^2 + 234y^2 - 66x - 255y + 63xy + 68 = 0,\\
\beta_5(x, y) &\cong -54x^2 + 336y^2 + 9x - 334y + 108xy + 65 = 0,\\
\beta_6(x, y) &\cong -6x^2 + 12y^2 + x - 16y + 12xy + 3 = 0.
\end{aligned}$$

In view of (3.1), (3.2) and remark 3.1, we are now set to write the basis functions we appeal to relations (5.1), (5.2), (5.4) and (5.6).

$$\begin{aligned}
W_1(a_1) &= \frac{H_1l_2l_3\alpha_{11}\alpha_{12}}{l^2}, & W_{11}(a_{11}) &= \frac{H_{11}l_1^*l_2l_3l_4\alpha_{12}}{l^2},\\
W_{12}(a_{12}) &= \frac{H_{12}l_1^*l_2l_3l_4\alpha_{11}}{l^2}, & W_2(a_2) &= \frac{H_2l_3l_4\alpha_{21}\alpha_{22}}{l^2},\\
W_{21}(a_{21}) &= \frac{H_{21}l_1l_2^*l_3l_4\alpha_{22}}{l^2}, & W_{22}(a_{22}) &= \frac{H_{22}l_1l_2^*l_3l_4\alpha_{21}}{l^2},\\
W_3(a_3) &= \frac{H_3l_1l_4\alpha_{31}\alpha_{32}}{l^2}, & W_{31}(a_{31}) &= \frac{H_{31}l_1l_2l_3^*l_4\alpha_{32}}{l^2},\\
W_{32}(a_{32}) &= \frac{H_{32}l_1l_2l_3^*l_4\alpha_{31}}{l^2}, & W_4(a_4) &= \frac{H_4l_1l_2\alpha_{41}\alpha_{42}}{l^2},\\
W_{41}(a_{41}) &= \frac{H_{41}l_1l_2l_3l_4^*\alpha_{42}}{l^2}, & W_{42}(a_{42}) &= \frac{H_{42}l_1l_2l_3l_4^*\alpha_{41}}{l^2}.
\end{aligned}$$

For convenience, let us denote by $B_4 = l_1l_2l_3l_4$, then we have

$$\begin{aligned}
W_1^*(b_1) &= \frac{H_1^*B_4\beta_1}{l^2}, & W_2^*(b_2) &= \frac{H_2^*B_4\beta_2}{l^2}, & W_3^*(b_3) &= \frac{H_3^*B_4\beta_3}{l^2},\\
W_4^*(b_4) &= \frac{H_4^*B_4\beta_4}{l^2}, & W_5^*(b_5) &= \frac{H_5^*B_4\beta_5}{l^2}, & W_6^*(b_6) &= \frac{H_6^*B_4\beta_6}{l^2},
\end{aligned}$$

where H_r^* $(r = 1, 2, \ldots, 6)$ are normalizing constants. It may be verified easily that t properties for wedges specified in article 1.5 of [8] hold for the above construction of t wedges.

Acknowledgement

The authors would like to thank the referees of this paper for their helpful suggestion

References

[1] Ciarlet Phillipe G, *The finite element method for elliptic problems* (Amsterdam, New Yorl Oxford: North-Holland Publishing Company) (1978)

[2] Coxeter H S M, *Introduction to geometry* (New York: Wiley) (1961)

[3] Gout J L, Construction of a Hermite rational 'Wachspress type' finite element, *Comp. Math. Appl.* **5** (1979) 337–347

[4] Oden J Tinsley and Carey Graham H, Finite Elements, A Second Course, (New Jersey: Prentice-Hall, Inc., Englewood Cliffs) (1983) vol. II

[5] Powar P L, C^1-approximation over triangles, in a Special Symposium on Approximation Theory and its Applications, edited by Geetha S Rao (New Age International Ltd. Publisher) (1996) 99–106

[6] Powar P L, Rana S S and Rao R, Local behaviour of the denominator in the construction of rational finite element basis over rectangles, communicated

[7] Powar P L and Rao R, A counter example of the construction of a C^1 rational finite element due to Wachspress, *Comp. Math. Appl.* **22** (1991) 17–22

[8] Wachspress E L, *A Rational Finite Element Basis* (New York: Academic Press) (1975)

Proc. Indian Acad. Sci. (Math. Sci.), Vol. 110, No. 1, February 2000, pp. 79–101.
© Printed in India

A direct heuristic algorithm for linear programming

S K SEN and A RAMFUL*

Supercomputer Education and Research Centre, Indian Institute of Science, Bangalore 560012, India
*Department of Mathematics, University of Mauritius, Reduit, Mauritius

MS received 13 August 1998

Abstract. An $O(n^3)$ mathematically non-iterative heuristic procedure that needs no artificial variable is presented for solving linear programming problems. An optimality test is included. Numerical experiments depict the utility/scope of such a procedure.

Keywords. Direct heuristic algorithm for linear programming; interior-point methods; optimality test; p-inverse; revised simplex algorithm.

1. Introduction

The simplex method [6, 23] – an exponential time (non-polynomial time) algorithm – or its variation has been used and is being used to solve almost any linear programming problem (LPP) for the last four decades. In 1979, Khachiyan proposed the ellipsoid method – the first polynomial-time (interior-point) algorithm – to solve LPPs [13]. Then, in 1984, Karmarkar suggested the second polynomial time $(O(n^{3.5}))$ algorithm based on projective transformation [11, 12, 22, 24]. Unlike the ellipsoid method, Karmarkar method appears to solve very large LPPs faster than does the simplex method. However, most of the available software packages that we know of for solving LPPs are still based on the simplex algorithm (SA) or a variation of it. Both the ellipsoid method and the Karmarkar method are mathematically iterative and need many times more computing resources than does the SA certainly for small LPPs and possibly for reasonably large LPPs. Consequently, none of these are so far popular nor are they known to be used commercially extensively like the SA.

There exists no mathematically direct algorithm to solve LPPs like the ones (e.g., Gauss reduction with partial pivoting) to solve linear systems. In fact, if we view the LPP geometrically it would not be far to see why it is difficult to have a non-iterative algorithm (where the exact number of arithmetic operations is known *a priori*).

The word iteration has one meaning in computer science and a different meaning in mathematics. For example, the multiplication of two $n \times n$ matrices $A = [a_{ij}]$ and $B = [b_{ij}]$ to get the matrix $C = [c_{ij}]$, where $c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$, in the usual way is iterative in computer science while it is non-iterative or, equivalently, direct (i.e., we know the exact number of operations to obtain c_{ij} beforehand) in mathematics. In fact, iteration means simply repetition in computer science.

Each linear equation in a linear system represents a hyperplane geometrically. The intersection of all the hyperplanes corresponding to the equations in the linear system will be a convex region. The portion of this region that falls [32, 19] in the first quadrant (i.e. in the non-negative region) is defined as a polytope.

In the LPP "minimise $\mathbf{c}^t\mathbf{x}$ (objective function) subject to $\mathbf{Ax} = \mathbf{b}$ (linear constraints), $\mathbf{x} \geq \mathbf{0}$ (non-negativity condition)", $\mathbf{Ax} = \mathbf{b}$ and $\mathbf{x} \geq \mathbf{0}$ (null column vector) define the polytope. One of the corners of the polytope is the required solution. We know *exactly* the direction of search, viz., the direction of the vector $\mathbf{c}$; but we do not know the point or location from which we should start in the direction of $\mathbf{c}$. If we already know this location then all LPPs can be solved mathematically directly just as we can solve linear equations directly. Since we do not have the knowledge of this location, we use some kind of trial and error procedure or a procedure which is implicitly a trial and error one to finally hit upon the optimal solution. For example, in the polytope-shrinking interior-point method [19] we start from a centre (middle) of the polytope (which appears quite reasonable) and proceed in the direction of $\mathbf{c}$. Since the centre does not happen to be the required location, we hit a hyperplane instead of the desired corner. This will help in deciding the next location by some means; for example, a hyperplane normal to $\mathbf{c}$ could be drawn from the point of hit such that the resulting (greatly shrunk) polytope above the hyperplane will be the one for the next search. We again start from a centre of the shrunk polytope and proceed in the direction of $\mathbf{c}$. We continue this process till either we hit the desired corner or determine this corner uniquely from the remaining hyperplanes by solving the equations corresponding to these hyperplanes.

Can we really solve LPPs directly in $O(n^3)$ operations just like the way we solve linear systems? The proposed procedure is essentially an attempt to answer this question. We have been able to find out a few problems where the procedure does not give the optimal solution. However, even if it does not, this heuristic procedure still gives a basic feasible solution quite close to the actual optimal solution with, however, a set of basic variables different from the actual ones. One can make use of this solution to obtain the optimal solution in a fewer iterations through the revised simplex procedure [35] or a variation of it.

In § 2, we describe the direct heuristic algorithm for linear programming (DHALP) and discuss a few results concerning the algorithm. We illustrate the procedure by numerical examples and state our observations including effectiveness of the procedure through numerical experiments in § 3, while in § 4 we include the conclusions and specifically demonstrate that the proposed heuristic algorithm is distinctly different from the popular SA not only in not considering the artificial variables but also in the detection of the basic variables deterministically. Also we compare the DHALP with interior-point [13, 19, 29, 36] and other methods including the inequality sorting algorithms [15, 17].

2. Direct heuristic algorithm

We first present here the LPP along with the DHALP without any comment or justification and then discuss a few results on the DHALP and on its computational/space complexity.

The LPP. Let the LPP be written in the form (without loss of generality)

$$\text{minimize } z = \mathbf{c}^t\mathbf{x} \text{ subject to } A\mathbf{x} = \mathbf{b}, \quad \mathbf{x} \geq \mathbf{0} \tag{1}$$

where $A = [a_{ij}]$ is an $m \times n$ constraint matrix, $\mathbf{c} = [c_i]$ is an $n \times 1$ column vector, $\mathbf{b} = [b_j]$ is an $m \times 1$ column vector, t indicates the transpose, and $\mathbf{0}$ represents the $n \times 1$ null column vector.

2.1 *The DHALP with optimality test*

The inputs are $A, \mathbf{b}, \mathbf{c}$ while the outputs are the $n \times 1$ solution vector $\mathbf{x} = [x_i]$, the value of the objective function z, and the comments based on the optimality test.

S.1: Input $m, n, A = [a_{ij}], i = 1(1)m; j = 1(1)n, b = [b_j],\ j = 1(1)m, \mathbf{c} = [c_i], i = 1(1)n$, where $i = 1(1)m$ implies $i = 1, 2, 3, \ldots, m$.

S.1a: Initialize indexarray by n zeros.

S.2: Compute $\mathbf{d} = A^+\mathbf{b}$, where $\mathbf{d} = [d_i]$ is an $n \times 1$ vector and A^+ is the Moore-Penrose inverse or, equivalently, minimum norm least square inverse or, equivalently, pseudoinverse or, equivalently, p-inverse [21, 25, 26, 10, 8, 34, 27, 28, 4, 30, 31, 5, 7, 20, 14, 16] $\mathbf{e} = A\mathbf{d}$.

If $\mathbf{e} \neq \mathbf{b}$ then output 'The LPP is inconsistent or, equivalently, infeasible.' Terminate.

S.3: Compute

$$\begin{aligned} H &= A^+A, \\ \mathbf{c}' &= (I - H)\mathbf{c}, \\ s_k &= \min\left\{\frac{d_i}{c_i'}; c_i' > 0\right\}, \\ \mathbf{x} &= \mathbf{d} - \mathbf{c}'s_k, \end{aligned}$$

where $\mathbf{c}' = [c_i']$ is an $n \times 1$ vector, I is an $n \times n$ unit matrix, H is an $n \times n$ matrix. The direction vector $\mathbf{c}'s_k$ attempts to push the point (the solution vector $\mathbf{x}$) of the solution space defined by $A\mathbf{x} = \mathbf{b}$ into the polytope defined by $A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}$ if it is not already in the polytope. The successive computation produces the point (the required solution vector $\mathbf{x}$ of the LPP) at one of the corners of the polytope.

S.4: Remove the x_i that becomes zero, remove the (corresponding) ith column vector of the constraint matrix A, and the (corresponding) element c_i from the $\mathbf{c}$-vector of the objective function, shrink A and $\mathbf{c}$ and maintain an index counter for the variable x_i that has been removed. Reduce the dimension n of A by one, n of $\mathbf{c}$ by one. Compute $\mathbf{d} = A^+\mathbf{b}$.

Remarks. The foregoing A^+ in the step S.4 is computed from the current (most recent) A^+. This computation takes $O(mn)$ operations. One could have computed A^+ from the shrunk A only, but this would have taken $O(mn^2)$ operations. If two or more x_i become zero then one has to execute the algorithm removing one x_i at a time and computing the corresponding final solution $\mathbf{x}$. The $\mathbf{x}$ that gives the minimum value of the objective function will be the required solution (output) of the DHALP. There is no easy way to decide which one of the two (or more) null x_i should leave the basis. However, such a situation is not too common.

S.5: Repeat the steps S.3 and S.4 till s_k is zero or not computable (i.e., $c_i' \leq 0$ for all i).

S.6: Compute the value of the objective function $z = \mathbf{c}^t\mathbf{x}$ where $\mathbf{x}$ and $\mathbf{c}$ are the most recent vectors and output the results.

The optimality test. Let the resulting solution vector $\mathbf{x}$ that we obtain from the DHALP, the quantities A, $\mathbf{b}$, and $\mathbf{c}$ of eq. (1) be known. Also, let B be the basis (columns of A corresponding to the basic variables x_i of the foregoing vector $\mathbf{x}$), $\mathbf{c}_B^t$ be the vector with the elements of $\mathbf{c}^t$ corresponding to the basic variables x_i, and $\mathbf{p}_j$ be the jth nonbasic column of A (i.e., the column of A whose coefficient x_j is a nonbasic variable).

S.7: Compute $\mathbf{y}^t = \mathbf{c}_B^t B^{-1}$ (a row vector).

S.8: Compute $z_j - c_j = \mathbf{y}^t \mathbf{p}_j - c_j$ (a scalar) for all nonbasic vectors $\mathbf{p}_j$.

S.9: Test the sign of $z_j - c_j$. If all $z_j - c_j \leq 0$ then the solution is optimal (minimal) else the solution is unbounded (no lower bound for the minimization problem) or the DHALP could not reach the optimal solution; use the revised SA [35] to obtain the optimal solution starting from the results of DHALP, i.e., using the DHALP solution as the initial basic feasible solution.

2.2 *Justification of the DHALP*

The step S.1 of the DHALP is simply the input step. The step S.2 checks the consistency of the constraint equation $A\mathbf{x} = \mathbf{b}$ before proceeding further. If the vector $\mathbf{e} = AA^+\mathbf{b} \neq \mathbf{b}$ then the equation $A\mathbf{x} = \mathbf{b}$ is inconsistent [28]. Hence the LPP is infeasible and we terminate the DHALP. A^+ will be termed as the p-inverse of A in the rest of the article. 'Why the term p-inverse?' is explained in the article by Lakshmikantham *et al* [16].

The general solution of the linear system $A\mathbf{x} = \mathbf{b}$, where the vector $\mathbf{b}$ may be zero or not, is $\mathbf{x} = A^+\mathbf{b} \pm P\mathbf{z}$, where $P = (I - A^+A)$ is the orthogonal projection operator that projects any arbitrary vector $\mathbf{z}$ orthogonally onto the null space of the matrix A. We are essentially computing a point (represented by a vector in an n-dimensional space) $\mathbf{c}'$ in the null space of the matrix A in the step S.3. If we write $\mathbf{x} = \mathbf{d} - \mathbf{c}' s_k = A^+\mathbf{b} - (I - A^+A)\mathbf{c}s_k$, we can easily see that the solution vector $\mathbf{x}$ is of the form $A^+\mathbf{b} - P\mathbf{z}$, where $\mathbf{c}s_k$ corresponds to the arbitrary column vector $\mathbf{z}$. The scalar s_k in the step S.3 is computed so as to (i) make one (or more) element x_i of $\mathbf{x}$ zero, i.e., to make x_i nonbasic and (ii) reduce the value of the objective function. The step S.3 has also the effect of pushing the solution vector $\mathbf{x}$ into the polytope [19, 32] if it is not already in it. Sometimes, rarely though, a true basic variable x_i may turn out to be zero and hence nonbasic. Under which necessary and sufficient condition does an actual basic variable become nonbasic or, equivalently, what would be the conditions/restrictions on A, $\mathbf{b}$ and $\mathbf{c}$ so that the DHALP becomes a regular direct algorithm? This is an open problem which needs to be explored.

In the step S.4 we remove the variable x_i whose value has become zero once for all – quite often this x_i is nonbasic; at least in our numerical experiment, in over 95% problems, x_i with zero has been nonbasic.

The step S.5 includes a stopping condition while the step S.6 is the output step in which we may include appropriate comments for degenerate, infeasible, unbounded (without a lower bound) or infinite solution cases.

The optimality test is carried out in the steps S.7, S.8, and S.9. Usually the basis B will be a nonsingular (square) matrix. Testing for the optimality of the solution is straightforward. If B is rectangular $(m < n)$ then we append one (or more) row to B and one (or more) corresponding element to $\mathbf{b}$ so that the resulting matrix B and resulting vector $\mathbf{b}$ do not change $\mathbf{x}$, and B turns out to be square nonsingular. We then carry out the optimality test as in steps S.7, S.8, and S.9.

If the test fails, i.e., if not all $z_j - c_j \leq 0$ then we may use this basis B, without wasting/throwing away the computation due to the DHALP, in the revised SA to get the optimal solution in a comparatively few steps.

2.3 *Complexity of the DHALP*

We present here the order of computational and space complexities for the DHALP. We need, in the step S.2, (i) $O(mn^2)$ operations to compute the p-inverse of A, i.e., A^+ for the

$m \times n$ constraint matrix A [31, 10, 14], (ii) $O(mn)$ operations to compute the $n \times 1$ vector $\mathbf{d} = A^+\mathbf{b}$ and (iii) $O(mn)$ operations to compute the $m \times 1$ vector $\mathbf{e} = A\mathbf{d}$.

In step S.3, we may compute the orthogonal projection operator $P = (I - H) = (I - A^+A)$ directly (without computing A^+ or A^+A) using the concise algorithm for linear systems [18, 32, 19] However, instead, we compute the $n \times n$ matrix $H = A^+A$ in $O(mn^2)$ operations, $\mathbf{c}' = (I - H)\mathbf{c}$ in $O(n^2)$ operations, the scalar s_k in $O(n)$ operations, and the $n \times 1$ solution vector $\mathbf{x} = \mathbf{d} - \mathbf{c}'s_k$ in $O(n)$ operations.

We need, in the step S.4, $O(mn)$ operations for the removal of one column of A, one element of $\mathbf{c}$, one dimension for the number of columns n of A, and for shrinking A and $\mathbf{c}$ including keeping an index counter for the variable $\mathbf{x}_i$ that has been removed. For computing $\mathbf{d}$ in this step we need not have to compute A^+ from the original shrunk matrix A; instead, we compute A^+ from the most recently computed A^+. Thus we need $O(mn)$ operations to compute $\mathbf{d}$.

While repeating step S.5, we need not have to compute the p-inverse A^+ of the shrunk matrix A (as stated in the foregoing paragraph), that needs $O(mn^2)$ operations. We compute, instead, the new A^+ from the most recently computed A^+ in $O(mn)$ operations by the procedure PISM:

The procedure PISM: Let the current matrix A and its known p-inverse A^+ be denoted by A_{k+1} and A^+_{k+1}, respectively where $A_k = A_{k+1}$ without the qth column. We know both A_{k+1} and A^+_{k+1} as well as the column number q. Also, let $A^+_{k+1-q} = A^+_{k+1}$ without the qth row, $\mathbf{a}_q = q$th column of A_{k+1}, and $\mathbf{b}^t_q = q$th row of A^+_{k+1}. Then

$$A^+_k = A^+_{k+1-q} + \frac{1}{r}(A^+_{k+1-q}\mathbf{a}_q)\mathbf{b}^t_q, \tag{2}$$

where

$$r = 1 - \mathbf{b}^t_q\mathbf{a}_q. \tag{3}$$

Here r needs $O(m)$ operations, $A^+_{k+1-q}\mathbf{a}_q$ needs $O(nm)$ operations. The foregoing addition takes $O(mn)$ operations. Hence, the computation of each successive shrunk A^+ needs $O(mn)$ operations. The validity of eq. (2) can be easily verified by working backwards in Greville's algorithm where the qth row of $A^+_{k+1} = \mathbf{b}^t_q$ is already known [10, 14].

The computational complexity of the optimality test is as follows. Step S.7 needs $O(mn)$ operations to compute $\mathbf{y}^t$ for a nondegenerate bounded LPP – here B^{-1} is just the final A^+. Otherwise, it will take $O(m^3)$ operations. The step S.8 on the other hand, requires just $O(m)$ operations to compute $z_j - c_j$.

So far as the space complexity is concerned, we mainly need mn locations to store the constraint matrix A, m and n locations to store $\mathbf{b}$ and $\mathbf{c}$, respectively. During the successive removal of the elements of $\mathbf{x}$, we will be shrinking A as well as $\mathbf{c}$. Although this will reduce the need for the storage locations, this reduction may not be significant. We also need storage space for the program corresponding to the DHALP, which is not significant. Hence the space complexity is $O(mn)$.

3. Examples

We illustrate the DHALP by considering a few typical numerical examples. We also present an example where the DHALP has given a solution close to the optimal one (but not the optimal one), and demonstrate how to arrive at the optimal solution starting from the output (solution) of the DHALP.

Example 1 (*Nondegenerate bounded solution*). The LPP in an inequality form is

$$\min z = -1.2x_1 - 1.4x_2 \quad \text{subject to (s.t.)}$$
$$40x_1 + 25x_2 \leq 1000$$
$$35x_1 + 28x_2 \leq 980$$
$$25x_1 + 35x_2 \leq 875$$
$$x_1, x_2 \geq 0$$

We write the LPP as the one with equality constraints as follows

$$\min z = \mathbf{c}^t\mathbf{x} \;\; \text{s.t.}$$
$$A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0},$$

where

$$\mathbf{c} = \begin{bmatrix} -1.2 \\ -1.4 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}, \quad A = \begin{bmatrix} 40 & 25 & 1 & 0 & 0 \\ 35 & 28 & 0 & 1 & 0 \\ 25 & 35 & 0 & 0 & 1 \end{bmatrix},$$

$$\mathbf{b} = \begin{bmatrix} 1000 \\ 980 \\ 875 \end{bmatrix}, \quad m = 3, n = 5.$$

S.1: The inputs are the numerical vectors **c** and **b**, and the numerical matrix A.

S.1a: indexarray $= [0\ 0\ 0\ 0\ 0]^t$.

$$\text{S.2: } \mathbf{d} = A^+\mathbf{b} = \overbrace{\begin{bmatrix} 0.0361 & 0.0130 & -0.0361 \\ -0.0296 & -0.0035 & 0.0524 \\ 0.2966 & -0.4340 & 0.1345 \\ -0.4340 & 0.6417 & -0.2035 \\ 0.1345 & -0.2035 & 0.0682 \end{bmatrix}}^{A^+} \overbrace{\begin{bmatrix} 1000 \\ 980 \\ 875 \end{bmatrix}}^{\mathbf{b}} = \overbrace{\begin{bmatrix} 17.2657 \\ 12.8164 \\ -11.0406 \\ 16.8388 \\ -5.2187 \end{bmatrix}}^{\mathbf{d}}.$$

$$\mathbf{e} = A\mathbf{d} = [1000\ 980\ 875]^t.$$

Here $\mathbf{e} = \mathbf{b}$. (Hence the equation $A\mathbf{x} = \mathbf{b}$ is consistent.)

$$\text{S.3: } H = A^+A = \overbrace{\begin{bmatrix} 0.9972 & 0.0030 & 0.0361 & 0.0130 & -0.0361 \\ 0.0030 & 0.9963 & -0.0296 & -0.0035 & 0.0524 \\ 0.0361 & -0.0296 & 0.2966 & -0.4340 & 0.1345 \\ 0.0130 & -0.0035 & -0.4340 & 0.6417 & -0.2035 \\ -0.0361 & 0.0524 & 0.1345 & -0.2035 & 0.0682 \end{bmatrix}}^{H}$$

$$\mathbf{c}' = (I - H)\mathbf{c} = [0.0009\ \ -0.0015\ \ 0.0018\ \ 0.0107\ \ 0.0300]^t$$

$$s_k = \min\left\{\frac{d_i}{c_i'} : c_i' > 0\right\} = \min\left\{\frac{17.2657}{0.0009}, \frac{-11.0406}{0.0018}, \frac{16.8388}{0.0107}, \frac{-5.2187}{0.0300}\right\}.$$

$$s_k = \min\{19486, -5997, 1567, -174\} = -5997.$$

$$\mathbf{x} = \mathbf{d} - \mathbf{c}' s_k = [22.5793 \ \ 3.8732 \ \ 0 \ \ 81.2767 \ \ 174.9568]'.$$

S.4: We remove x_3 since it has become zero, i.e., nonbasic. Hence we remove the third column vector of A and the third element of $\mathbf{c}$, i.e., c_3. We then shrink the 3×5 matrix A and the 5×1 vector $\mathbf{c}$ to the 3×4 matrix and 4×1 vector and call them once again A and $\mathbf{c}$, respectively. The indexarray that keeps track of which element of $\mathbf{x}$ has become 0, i.e., nonbasic, now becomes $[0\ 0\ 3\ 0\ 0]'$. Replace n by $n-1$, i.e., n is now 4.

$$\mathbf{d} = A^+\mathbf{b} = [16.6990 \ \ 13.2815 \ \ 23.6506 \ \ -7.3298]'.$$

S.5: Now we go back to the step S.3.

$$\text{S.3: } H = \begin{bmatrix} 0.9991 & 0.0015 & -0.0092 & -0.0292 \\ 0.0015 & 0.9976 & 0.0148 & 0.0468 \\ -0.0092 & 0.0148 & 0.9094 & -0.2864 \\ -0.0292 & 0.0468 & -0.2864 & 0.0939 \end{bmatrix}$$

We compute the new A^+ from the current A^+ in $O(mn)$ operations using the procedure PISM as follows. Let the current matrix A be denoted by A_{k+1}, i.e.,

$$A_{k+1} = \begin{bmatrix} 40 & 25 & 1 & 0 & 0 \\ 35 & 28 & 0 & 1 & 0 \\ 25 & 35 & 0 & 0 & 1 \end{bmatrix}.$$

Let the current p-inverse of A be denoted by A_{k+1}^+, i.e.,

$$A_{k+1}^+ = \begin{bmatrix} 0.0361 & 0.0130 & -0.0361 \\ -0.0291 & -0.0035 & 0.0524 \\ 0.2966 & -0.4340 & 0.1345 \\ -0.4340 & 0.6417 & -0.2035 \\ 0.1345 & -0.2035 & 0.0682 \end{bmatrix}.$$

Let $A_k = A_{k+1}$ without the qth column. Here $q = 3$. So we remove the third column of the original matrix A. Hence

$$A_k = \begin{bmatrix} 40 & 25 & 0 & 0 \\ 35 & 28 & 1 & 0 \\ 25 & 35 & 0 & 1 \end{bmatrix}.$$

Moreover, let $A_{k+1-3}^+ = A_{k+1}^+$ without the third row, i.e.,

$$A_{k+1-3}^+ = \begin{bmatrix} 0.0361 & 0.0130 & -0.0361 \\ -0.0291 & -0.0035 & 0.0524 \\ -0.4340 & 0.6417 & -0.2035 \\ 0.1345 & -0.2035 & 0.0682 \end{bmatrix}.$$

Let $\mathbf{a}_3$ = third column of A_{k+1}, i.e., $\mathbf{a}_3 = [1 \ \ 0 \ \ 0]'$. Let $\mathbf{b}_3'$ = third row of A_{k+1}^+, i.e., $\mathbf{b}_3' = [0.2966 \ \ -0.4340 \ \ 0.1345]$. Compute $r = 1 - \mathbf{b}_3'\mathbf{a}_3 = 0.7034$. Compute

$$A_k^+ = A_{k+1-3}^+ + \frac{1}{r}(A_{k+1-3}^+\mathbf{a}_3)\mathbf{b}_3^t = \begin{bmatrix} 0.0513 & -0.0092 & -0.0292 \\ -0.0421 & 0.0148 & 0.0468 \\ -0.6170 & 0.9094 & -0.2864 \\ 0.1912 & -0.2864 & 0.0939 \end{bmatrix}.$$

$$\mathbf{c}' = [0.0010 \ \ -0.0016 \ \ 0.0096 \ \ 0.0304]^t, s_k = -241.1290.$$

$$\mathbf{x} = [16.9355 \ \ 12.9032 \ \ 25.9677 \ \ 0]^t.$$

We remove the last element of $\mathbf{x}$ since it has become zero. Hence we remove the la column of A and the last element of $\mathbf{c}$. We then shrink the 3×4 matrix A and the $4 \times$ vector $\mathbf{c}$ to the 3×3 matrix and 3×1 vector and call them once again A and respectively. The indexarray now becomes $[0 \ 0 \ 3 \ 0 \ 5]^t$ which means that the elements and x_5 have become nonbasic. Replace n by $n-1$, i.e., n is now 3.

S.5: We now go back to step S.3.

S.3: $H = A^+A$ is the unit matrix of order 3.
$\mathbf{c}' = (I - H)\mathbf{c} = \mathbf{0}$ (null column vector), s_k is not computable and hence the curre solution vector $\mathbf{x}$ using the information of the indexarray is

$$\mathbf{x} = [16.9355 \ \ 12.9032 \ \ 0 \ \ 25.9677 \ \ 0]^t.$$

S.6: The value of the objective function $z = \mathbf{c}^t\mathbf{x} = [-1.2 \ \ -1.4 \ \ 0 \ \ 0 \ \ 0]\mathbf{x} = -38.387$

Optimality test for the solution x: Here the basis (i.e., the original matrix A witho columns 3 and 5)

$$B = \begin{bmatrix} 40 & 25 & 0 \\ 35 & 28 & 1 \\ 25 & 35 & 0 \end{bmatrix}, \quad \mathbf{c}_B^t = \begin{bmatrix} -1.2 \\ -1.4 \\ 0 \end{bmatrix}, \quad c_3 = 0, \mathbf{p}_3 = \text{column 3 of origina}$$

$$A = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{p}_5 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

where $\mathbf{p}_j$ = column j of the original matrix A. Since B is nonsingular, $B^+ = B^{-1}$ is alrea available from step S.3.

S.7: $\mathbf{y}^t = \mathbf{c}_B^t B^{-1} = [-0.009 \ \ 0 \ \ -0.0335]$.

S.8: Compute $z_3 - c_3 = \mathbf{y}^t\mathbf{p}_3 - c_3 = -0.009$ since $c_3 = 0$. Similarly,

$$z_5 - c_5 = \mathbf{y}^t\mathbf{p}_5 - c_5 = -0.0335.$$

S.9: All (here two) the $z_j - c_j$ values are negative, i.e., ≤ 0. Hence the DHALP has giv us the optimal solution.

Example 2 (*Infeasible LPP where* $\boldsymbol{Ax} = \boldsymbol{b}$ *is inconsistent*)

$$\min z = x_1 + x_2 + x_3 \ \text{ s.t.}$$

$$5x_1 + 3x_2 + 2x_3 = 10$$

$$2x_1 + x_2 + 2x_3 = 5$$

$$4x_1 + 2x_2 + 4x_3 = 4$$
$$x_1, x_2, x_3 \geq 0$$

S.1: Input $m = 3, n = 3, A = \begin{bmatrix} 5 & 3 & 2 \\ 2 & 1 & 2 \\ 4 & 2 & 4 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 10 \\ 5 \\ 4 \end{bmatrix}$, $\mathbf{c} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

S.1.a: indexarray $= [0\ \ 0\ \ 0]^t$.

S.2: $\mathbf{d} = A^+\mathbf{b} = [1.6340\ \ 1.2491\ \ -0.9585]^t, \mathbf{e} = A\mathbf{d} = [10\ \ 2.6\ \ 5.2]^t \neq \mathbf{b}$. Output 'The LPP is inconsistent.' Terminate.

Example 3 (*Infeasible LPP where* $A\mathbf{x} = \mathbf{b}$ *is consistent*)

$$\min z = -20x_1 - 30x_2 \text{ s.t.}$$
$$2x_1 + 3x_2 \geq 120$$
$$x_1 + x_2 \leq 40$$
$$2x_1 + 1.5x_2 \geq 90$$
$$x_1, x_2 \geq 0$$

S.1: Input $m = 3, n = 5, A = \begin{bmatrix} 2 & 3 & -1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 2 & 1.5 & 0 & 0 & -1 \end{bmatrix}$, $\mathbf{b} = [120\ \ 40\ \ 90]^t$, $\mathbf{c} = [-20\ \ -30\ \ 0\ \ 0\ \ 0]^t$.

S.1a: indexarray $= [0\ \ 0\ \ 0\ \ 0\ \ 0]^t$.

S.2: $\mathbf{d} = A^+\mathbf{b} = [22.9231\ \ 23.0769\ \ -4.9231\ \ -6.0000\ \ -9.5385]^t$. $\mathbf{e} = A\mathbf{d} = [120\ \ 40\ \ 90]^t = \mathbf{b}$. Hence the equation $A\mathbf{x} = \mathbf{b}$ is consistent.

S.3: $\mathbf{c}' = (I - H)\mathbf{c} = [1.0769\ \ -3.0769\ \ -7.0769\ \ 2.0000\ \ -2.4615]^t, s_k = -3$. $\mathbf{x} = [x_1\ x_2\ x_3\ x_4\ x_5]^t = [26.1538\ \ 13.8462\ \ -26.1538\ \ 0\ \ -16.9231]^t$.

S.4: We remove x_4 since it has become zero. Hence we remove the fourth column vector of A and the fourth element of $\mathbf{c}$. Indexarray $= [0\ \ 0\ \ 0\ \ 4\ \ 0]^t$.

$$\mathbf{d} = A^+\mathbf{b} = [16.9231\ 23.0769\ \ -16.9231\ \ -21.5385]^t.$$

S.5: We now go back to the step S.3.

S.3: The matrix A^+ is computed using the procedure PISM in $O(mn)$ operations since the current A^+ and A are known and can be used.

$$\mathbf{c}' = (I - H)\mathbf{c} = [3.0769\ \ -3.0769\ \ -3.0769\ \ 1.5385]^t,\ \ s_k = -14.0000.$$
$$\mathbf{x} = [x_1\ x_2\ x_3\ x_5]^t = [60\ \ -20\ \ -60\ \ 0]^t.$$

S.4: We remove x_5 since it has become zero. Hence we remove the last column vector of A and the last element of $\mathbf{c}$. Indexarray $= [0\ \ 0\ \ 0\ \ 4\ \ 5]^t$.

$$\mathbf{d} = A^+\mathbf{b} = [60\ \ -20\ \ -60]^t.$$

S.5: We now go back to the step S.3.

S.3: $H = A^+A$ is the unit matrix of order 3. $\mathbf{c}' = (I - H)\mathbf{c} = [0\ \ 0\ \ 0]^t$.

Since $c'_i \le 0$ for all i, we cannot compute s_k. The final solution is $\mathbf{x} = [x_1\ \ x_2\ \ x_3\ \ x_4\ \ x_5]^t$ $[60\ \ -20\ \ -60\ \ 0\ \ 0]^t$ that contains two negative elements, viz., $x_2 = -20$ a $x_3 = -60$, violating the nonnegativity condition. Hence the problem is infeasible.

Remark. It may be seen in the foregoing minimization (infeasible) problem that the val of the objective function increased; in a feasible case the value decreases for ea successive removal of an element of **x**.

Example 4 (*Infeasible LPP with* $A\mathbf{x} = \mathbf{b}$ *consistent*)

$$\min z = -3x_1 - 2x_2 \text{ s.t.}$$

$$2x_1 + x_2 \le 2$$

$$3x_1 + 4x_2 \ge 12$$

$$x_1, x_2 \ge 0$$

S.1: Input $m = 2, n = 4, A = \begin{bmatrix} 2 & 1 & 1 & 0 \\ 3 & 4 & 0 & -1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 2 \\ 12 \end{bmatrix}$, $\mathbf{c} = [-3\ \ -2\ \ 0\ \ 0]^t$.

S.1.a: Indexarray $= [0\ \ 0\ \ 0\ \ 0]^t$.

S.2: $\mathbf{d} = A^+\mathbf{b} = [0.3571\ 2.5000\ -1.2143\ -0.9286]^t$. $\mathbf{e} = A\mathbf{d} = [2\ \ 12]^t = \mathbf{b}$. Hen the equation $A\mathbf{x} = \mathbf{b}$ is consistent.

S.3: $\mathbf{c}' = (I - H)\mathbf{c} = [-0.4643\ 0.2500\ 0.6786\ -0.3929]^t, s_k = -1.7895$.

$$\mathbf{x} = [x_1\ \ x_2\ \ x_3\ \ x_4]^t = [-0.4737\ \ 2.9474\ \ 0\ \ -1.6316]^t.$$

S.4: We remove x_3 since it has become zero. Hence we remove the third column vec of A and the third element of **c**.
Indexarray $= [0\ \ 0\ \ 3\ \ 0]^t, \mathbf{d} = A^+\mathbf{b} = [-0.5333\ \ 3.0667\ \ -1.3333]^t$.

S.5: We now go back to step S.3.

S.3: $\mathbf{c}' = (I - H)\mathbf{c} = [0.0333\ \ -0.0667\ \ -0.1667]^t$, $s_k = -16.0000$.

$$\mathbf{x} = [x_1\ \ x_2\ \ x_4]^t = [0\ \ 2\ \ -4]^t.$$

S.4: We remove x_1 since it has become zero. Hence we remove the first column vector A and the first element of **c**. indexarray $= [1\ \ 0\ \ 3\ \ 0]^t, \mathbf{d} = A^+\mathbf{b} = [2\ \ -4]^t$.

S.5: We now go back to step S.3.

S.3: $H = A^+A$ is now the unit matrix of order 2. $\mathbf{c}' = (I - H)\mathbf{c} = [0\ \ 0]^t$. Since $c'_i \le 0$ all i, we cannot compute s_k. The final solution is $\mathbf{x} = [x_1\ \ x_2\ \ x_3\ \ x_4]^t = [0\ \ 2\ \ 0\ \ -4]^t$ t contains one negative element, viz., $x_4 = -4$, violating the nonnegativity conditi Hence the problem is infeasible. As in the foregoing example, here also the value of objective function increased.

Example 5 (*Degenerate solution*)

$$\min z = -3x_1 - 9x_2 \text{ s.t.}$$

$$x_1 + 4x_2 \le 8$$

S.1: Input $m = 2, n = 4, A = \begin{bmatrix} 1 & 4 & 1 & 0 \\ 1 & 2 & 0 & 1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$, $\mathbf{c} = \begin{bmatrix} -3 \\ -9 \\ 0 \\ 0 \end{bmatrix}$.

S.1a: Indexarray $= [0\ 0\ 0\ 0\ 0]^t$.

S.2: $\mathbf{d} = A^+\mathbf{b} = [0.4444\ \ 1.7778\ \ 0.4444\ \ 0]^t \cdot \mathbf{e} = A\mathbf{d} = [8\ 4]^t = \mathbf{b}$. Hence the equation $A\mathbf{x} = \mathbf{b}$ is consistent.

S.3: $\mathbf{c}' = (I - H)\mathbf{c} = [-0.3333\ \ -0.3333\ \ 1.6667\ \ 1.0000]^t$, $s_k = 0$.

$$\mathbf{x} = [x_1\ x_2\ x_3\ x_4]^t = [0.4444\ \ 1.7778\ \ 0.4444\ 0]^t.$$

S.4: We now remove x_4 since it has become zero. Hence we remove the fourth column vector of A and fourth element of $\mathbf{c}$. Indexarray $= [0\ 0\ 0\ 4]^t$.
$\mathbf{d} = A^+\mathbf{b} = [0.4444\ \ 1.7778\ \ 0.4444]^t$.

S.5: We now go back to step S.3.

S.3: $\mathbf{c}' = (I - H)\mathbf{c} = [0.6667\ \ -0.3333\ \ 0.6667]^t$, $s_k = 0.6667$.

$$\mathbf{x} = [x_1\ x_2\ x_3]^t = [0\ 2\ 0]^t.$$

S.4: Since $\mathbf{x}$ contains two zeros, we have two possibilities. (i) Keep x_1 in the basis and remove x_3 and (ii) keep x_3 in the basis and remove x_1.

(i) We remove x_3 from the basis. Thus indexarray $= [0\ 0\ 3\ 4]^t$. $\mathbf{d} = A^+\mathbf{b} = [0\ 2]^t$.

S.5: We now go back to step S.3.

S.3: $H = A^+A$ is the unit matrix of order 2. $\mathbf{c}' = (I - H)\mathbf{c} = [0\ 0]^t$. Since $c'_i \leq 0$ for all i, we cannot compute s_k.

S.6: Compute $\mathbf{z} = \mathbf{c}^t\mathbf{x} = (-3 \times 0) + (-9 \times 2) = -18$. Output

$$\mathbf{x} = [x_1\ x_2\ x_3\ x_4]^t = [0\ 2\ 0\ 0]^t.$$

We subject the solution (basic) to the optimality test. Here

$$B = \begin{bmatrix} 1 & 4 \\ 1 & 2 \end{bmatrix}, \quad \mathbf{c}_B = [-3 - 9]^t, \quad \mathbf{p}_3 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \mathbf{p}_4 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad c_3 = 0, \quad c_4 = 0.$$

S.7: Compute $\mathbf{y}^t = \mathbf{c}_B^t B^{-1} = [-1.5\ \ -1.5]^t$.

S.8: Compute $z_3 - c_3 = \mathbf{y}^t\mathbf{p}_3 - c_3 = -1.5$, and $z_5 - c_5 = \mathbf{y}^t\mathbf{p}_5 - c_5 = -1.5$.

S.9: As all the $z_j - c_j$ values are negative, the DHALP has given us the optimal solution to the degenerate problem with the basic variable $x_1 = 0$. (ii) We remove x_1 from the basis. Indexarray $= [1\ \ 0\ \ 0\ \ 4]^t, \mathbf{d} = A^+\mathbf{b} = [2\ \ 0]^t$.

S.5: We now go back to step S.3.

S.3: $H = A^+A$ is the unit matrix of order 2. $\mathbf{c}' = (I - H)\mathbf{c} = [0\ \ 0]^t$. Since $c'_i \leq 0$ for all i, we cannot compute s_k.

S.6: Output $z = \mathbf{c}^t\mathbf{x} = (-9 \times 2) + (0 \times 0) = -18$ and

$$\mathbf{x} = [x_1\ x_2\ x_3\ x_4]^t = [0\ 2\ 0\ 0]^t.$$

We subject the solution (basic) to the optimality test. Here

$$B = \begin{bmatrix} 4 & 1 \\ 2 & 0 \end{bmatrix}, \quad \mathbf{c}_B = [-9\ \ 0]^t, \quad \mathbf{p}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{p}_4 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad c_1 = -3, \quad c_4 = 0.$$

S.7: Compute $\mathbf{y}^t = \mathbf{c}_B^t B^{-1} = [0 \;\; -4.5]^t$.

S.8: Compute $z_1 - c_1 = \mathbf{y}^t\mathbf{p}_1 - c_1 = -1.5$, and $z_4 - c_4 = \mathbf{y}^t\mathbf{p}_4 - c_4 = -4.5$.

S.9: As all the $z_j - c_j$ values are negative, the DHALP has given us the optimal solutic to the degenerate problem with the basic variable $x_3 = 0$.

Optimal solution $\mathbf{x} = [0 \;\; 2 \;\; 0 \;\; 0]^t, z = -18$.

Example 6 (*Infinity of solutions*)

$$\min z = -2x_1 - 4x_2 \;\; \text{s.t.}$$
$$x_1 + 2x_2 \leq 5$$
$$x_1 + x_2 \leq 4$$
$$x_1, x_2 \geq 0$$

S.1: Input $m = 2, n = 4, A = \begin{bmatrix} 1 & 2 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 5 \\ 4 \end{bmatrix}$, $\mathbf{c} = [-2 \;\; -4 \;\; 0 \;\; 0]^t$.

S.1a: Indexarray $= [0 \;\; 0 \;\; 0 \;\; 0]^t$.

S.2: $\mathbf{d} = A^+\mathbf{b} = [1.3333 \;\; 1.6667 \;\; 0.3333 \;\; 1.0000]^t$. $\mathbf{e} = A\mathbf{d} = [5 \;\; 4]^t = \mathbf{b}$. Hence tl equation $A\mathbf{x} = \mathbf{b}$ is consistent.

S.3: $\mathbf{c}' = (I - H)\mathbf{c} = [0.0000 \;\; -0.6667 \;\; 1.3333 \;\; 0.6667]^t, s_k = 0.25$.

$$\mathbf{x} = [x_1 \;\; x_2 \;\; x_3 \;\; x_4]^t = [1.3333 \;\; 1.8333 \;\; 0.0000 \;\; 0.8333]^t.$$

S.4: We remove x_3 since it has become zero. Hence we remove the third column vect of A and third element of $\mathbf{c}$. Indexarray $= [0 \;\; 0 \;\; 3 \;\; 0]^t$.

$$\mathbf{d} = A^+\mathbf{b} = [1.3333 \;\; 1.8333 \;\; 0.8333]^t.$$

S.5: We now go back to the step S.3.

S.3: $\mathbf{c}' = (I - H)\mathbf{c} = [0 \;\; 0 \;\; 0]^t$. Since $c_i' \leq 0$ for all i, s_k is not computable.

S.6: Compute $z = \mathbf{c}^t\mathbf{x} = (-2 \times 1.3333) + (-4 \times 1.8333) + (0 \times 0.8333) = -10$. Ou put $\mathbf{x} = [x_1 \;\; x_2 \;\; x_3 \;\; x_4]^t = [1.3333 \;\; 1.8333 \;\; 0.0000 \;\; 0.8333]^t$, $z = -10$. We subject t solution (basic) to the optimality test. Here

$$B = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 1 & 1 \end{bmatrix}, \quad \mathbf{c}_B = \begin{bmatrix} -2 \\ -4 \\ 0 \end{bmatrix}, \quad \mathbf{p}_3 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad c_3 = 0.$$

When we have the number of equations less than the number of variables and the soluti is bounded, we append additional rows to the rectangular basis matrix B so that t resulting B is nonsingular (square). The corresponding righthand side value of the n element of $\mathbf{b}$ is then computed using the solution vector $\mathbf{x}$. The columns $\mathbf{p}_j$ (of corresponding to the nonbasic variables x_j are appended by zeros so that their dimensi is compatible with the order of the basis B (i.e., the number of elements in $\mathbf{p}_j$ = the orc of B). This appendage (by zeros) is equivalent to appending null rows to A a corresponding zero elements to $\mathbf{b}$. The optimality test is then performed.

The basis matrix $B = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 1 & 1 \end{bmatrix}$ with the appended row $[0 \;\; 0 \;\; 1]$ and the nonba vector $\mathbf{p}_3$ with one appended zero become

$$B = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{p}_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

S.7: Compute $\mathbf{y}^t = \mathbf{c}_B^t B^{-1} = [-2 \;\; 4 \;\; 0]B^{-1} = [-2 \;\; 0 \;\; 0]$.

S.8: Compute $z_3 - c_3 = \mathbf{y}^t \mathbf{p}_3 - c_3 = [-2 \;\; 0 \;\; 0]\mathbf{p}_3 - 0 = -2$.

S.9: As $z_3 - c_3 \leq 0$ corresponding to the only nonbasic variable x_3, the DHALP has given us the optimal solution.

Example 7 (*Unbounded case*)

$$\begin{aligned} &\min z = -100x_1 - 800x_2 \;\; \text{s.t.} \\ &6x_1 + 2x_2 \geq 12 \\ &2x_1 + 2x_2 \geq 8 \\ &4x_1 + 12x_2 \geq 24 \\ &x_1, x_2 \geq 0 \end{aligned}$$

S.1: Input $m = 3, n = 5$,

$$A = \begin{bmatrix} 6 & 2 & -1 & 0 & 0 \\ 2 & 2 & 0 & -1 & 0 \\ 4 & 12 & 0 & 0 & -1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 12 \\ 8 \\ 24 \end{bmatrix}, \quad \mathbf{c} = [-100 \; - 800 \; 0 \; 0 \; 0]^t.$$

S.1a: Indexarray $= [0 \;\; 0 \;\; 0 \;\; 0 \;\; 0]^t$.

S.2: $\mathbf{d} = A^+\mathbf{b} = [1.5481 \;\; 1.4962 \;\; 0.2811 \;\; -1.9114 \;\; 0.1470]^t$.
$\mathbf{e} = A\mathbf{d} = [12 \;\; 8 \;\; 24]^t = \mathbf{b}$. Hence the equation $A\mathbf{x} = \mathbf{b}$ is consistent.

S.3: $\mathbf{c}' = (I - H)\mathbf{c} = [7.7622 \;\; -8.4757 \;\; 29.6216 \;\; -1.4270 \;\; -70.6595]^t$, $s_k = 0.0095$.

$$\mathbf{x} = [x_1 \;\; x_2 \;\; x_3 \;\; x_4 \;\; x_5]^t = [1.4745 \;\; 1.5766 \;\; 0 \;\; -1.8978 \;\; 0.8175]^t.$$

S.4: We remove x_3 since it has become zero. Hence we remove the third column vector of A and the third element of $\mathbf{c}$. Indexarray $= [0 \;\; 0 \;\; 3 \;\; 0 \;\; 0]^t$.

$$\mathbf{d} = A^+\mathbf{b} = [1.4952 \;\; 1.5143 \;\; -1.9810 \;\; 0.1524]^t.$$

S.5: We now go back to the step S.3.

S.3: $\mathbf{c}' = (I - H)\mathbf{c} = [2.1905 \;\; -6.5714 \;\; -8.7619 \;\; -70.0952]^t$, $s_k = 0.6826$.

$$\mathbf{x} = [x_1 \;\; x_2 \;\; x_4 \;\; x_5]^t = [0.0000 \;\; 6.0000 \;\; 4.0000 \;\; 48.0000]^t.$$

S.4: We remove x_1 since it has become zero. Hence we remove the first column vector of A and the first element of $\mathbf{c}$. indexarray $= [1 \;\; 0 \;\; 3 \;\; 0 \;\; 0]^t$.

$$\mathbf{d} = A^+\mathbf{b} = [6.0000 \;\; 4.0000 \;\; 48.0000]^t.$$

S.5: We now go back to step S.3.

S.3: $H = A^+A$ is the unit matrix of order 3. Sincc $c_i' \leq 0$ for all i, we cannot compute s_k.

S.6: Compute $z = \mathbf{c}^t\mathbf{x} = (-800 \times 6) + (0 \times 4) + (0 \times 48) = -4800$.

Output $\mathbf{x} = [x_1 \;\; x_2 \;\; x_3 \;\; x_4 \;\; x_5]^t = [0 \;\; 6 \;\; 0 \;\; 4 \;\; 48]^t, z = -4800$.

We subject the solution (basic) to the optimality test. Here

$$B=\begin{bmatrix}2 & 0 & 0\\ 2 & -1 & 0\\ 12 & 0 & -1\end{bmatrix},\quad \mathbf{c}_B=\begin{bmatrix}-800\\ 0\\ 0\end{bmatrix},\quad \mathbf{p}_1=\begin{bmatrix}6\\ 2\\ 4\end{bmatrix},$$

$$\mathbf{p}_3=\begin{bmatrix}-1\\ 0\\ 0\end{bmatrix},\quad c_1=-100, c_3=0.$$

S.7: Compute $\mathbf{y}^t=\mathbf{c}_B^tB^{-1}=[-400\ \ 0\ \ 0]^t$.

S.8: Compute $z_1-c_1=\mathbf{y}^t\mathbf{p}_1-c_1=-2400$ and $z_3-c_3=\mathbf{y}^t\mathbf{p}_3-c_3=400$.

S.9: Since $z_3-c_3>0$, the optimal solution is not reached. Either the solution is unbounded or it is not optimal. We use the revised SA to proceed further, starting with the output solution $\mathbf{x}$ of the DHALP algorithm as the initial basic feasible solution.

Revised SA: From the DHALP, the nonbasic variables are x_1 and x_3.

S.10: *Determining the entering vector* $\mathbf{p}_j$: Compute z_j-c_j for nonbasic $\mathbf{p}_1$ and $\mathbf{p}_3$. From the step S.8,

$z_1-c_1=\mathbf{y}^t\mathbf{p}_1-c_1=-2400$ and $z_3-c_3=\mathbf{y}^t\mathbf{p}_3-c_3=400$. Since z_3-c_3 is positive, x_3 now becomes a basic variable and $\mathbf{p}_3$ enters the basis.

S.11: *Determining the leaving vector* $\mathbf{p}_j$

(i) The current basic solution is

$$\mathbf{x}_B=\begin{bmatrix}x_2\\ x_4\\ x_5\end{bmatrix}=B^{-1}\mathbf{b}=\begin{bmatrix}0.5 & 0 & 0\\ 1 & -1 & 0\\ 6 & 0 & -1\end{bmatrix}\begin{bmatrix}12\\ 8\\ 4\end{bmatrix}=\begin{bmatrix}6\\ 4\\ 48\end{bmatrix}.$$

(ii) The constraint coefficients of the entering variables, i.e.,

$$\alpha^{(3)}=B^{-1}\mathbf{p}_3=\begin{bmatrix}0.5 & 0 & 0\\ 1 & -1 & 0\\ 6 & 0 & -1\end{bmatrix}\begin{bmatrix}-1\\ 0\\ 0\end{bmatrix}=\begin{bmatrix}-0.5\\ -1\\ -6\end{bmatrix}.$$

$$\theta=\min_k\left\{\frac{(B^{-1}\mathbf{b})_k}{\alpha_k^{(3)}},\alpha_k^{(3)}>0\right\}$$

where $(B^{-1}\mathbf{b})_k$ and $\alpha_k^{(3)}$ are the kth element of $B^{-1}\mathbf{b}$ and $\boldsymbol{\alpha}^{(3)}$. Since all the $\alpha_k^{(3)}\le 0$, the problem has no bounded solution.

Example 8 (*Cycling in SA*). The LPP [3, 1] is

$$\min z=-0.75x_4+20x_5-0.5x_6+6x_7 \text{ s.t.}$$

$$x_1+0.25x_4-8x_5-x_6+9x_7=0$$

$$x_2+0.5x_4-12x_5-0.5x_6+3x_7=0$$

$$x_3+x_6=1$$

$$x_1,x_2,x_3,x_4,x_5,x_6,x_7\ge 0$$

S.1: Input $m = 3, n = 7$,

$$A = \begin{bmatrix} 1.0000 & 0 & 0 & 0.2500 & -8.0000 & -1.0000 & 9.0000 \\ 0 & 1.0000 & 0 & 0.5000 & -12.0000 & -0.5000 & 3.0000 \\ 0 & 0 & 1.0000 & 0 & 0 & 1.0000 & 0 \end{bmatrix},$$

$$\mathbf{b} = [0\ 0\ 1]^t, \quad \mathbf{c} = [0\ 0\ 0\ -0.75\ 20\ -0.5\ 6]^t.$$

S.1a: Indexarray $= [0\ 0\ 0\ 0\ 0\ 0\ 0]^t$.

S.2: $\mathbf{d} = A^+\mathbf{b} = [0.0063\ -0.0034\ 0.5023\ -0.0001\ -0.0095\ 0.4977\ 0.0462]^t$. $\mathbf{e} = A\mathbf{d} = [0\ 0\ 1]^t = \mathbf{b}$. Hence the equation $A\mathbf{x} = \mathbf{b}$ is consistent.

S.3: $\mathbf{c}' = (I - H)\mathbf{c} = [-1.4946\ 2.6341\ 0.1612\ 0.1934\ 0.3471\ -0.1612\ 0.4513]^t$, $s_k = -0.0273$, $\mathbf{x} = [x_1\ x_2\ x_3\ x_4\ x_5\ x_6\ x_7]^t = [-0.0346\ 0.0686\ 0.5067\ 0.0052\ 0\ 0.4933\ 0.0585]^t$.

S.4: We remove x_5 since it has become zero. Hence we remove the fifth column vector of A and the fifth element of $\mathbf{c}$. indexarray $= [0\ 0\ 0\ 0\ 5\ 0\ 0]^t$.

$$\mathbf{d} = A^+\mathbf{b} = [-0.0155\ 0.0649\ 0.5085\ 0.0286\ 0.4915\ 0.0555]^t.$$

S.5: We now go back to step S.3.

S.3: $\mathbf{c}' = (I - H)\mathbf{c} = [-0.6996\ 0.1350\ -0.0660\ -0.8570\ 0.0660\ 0.1089]^t, s_k = 0.4803$.

$$\mathbf{x} = [x_1\ x_2\ x_3\ x_4\ x_6\ x_7]^t = [-0.3206\ 0\ 0.5402\ 0.4404\ 0.4598\ 0.0032]^t.$$

S.4: We remove x_2 since it has become zero. Hence we remove the second column vector of A and the second element of $\mathbf{c}$. Indexarray $= [0\ 2\ 0\ 0\ 5\ 0\ 0]^t$.

$$\mathbf{d} = A^+\mathbf{b} = [-0.0875\ 0.5268\ 0.1193\ 0.4732\ 0.0590]^t.$$

S.5: We now go back to the step S.3.

S.3: $\mathbf{c}' = (I - H)\mathbf{c} = [-0.8497\ -0.0279\ -0.6686\ 0.0279\ 0.1161]^t$, $s_k = 0.5082$.

$$\mathbf{x} = [x_1\ x_3\ x_4\ x_6\ x_7]^t = [0.3443\ 0.5410\ 0.4590\ 0.4590\ 0]^t.$$

S.4: We remove x_7 since it has become zero. Hence we remove the last column vector of A and the last element of $\mathbf{c}$. indexarray $= [0\ 2\ 0\ 0\ 5\ 0\ 7]^t$.

$$\mathbf{d} = A^+\,\mathbf{d} = [0.2105\ 0.7193\ 0.2807\ 0.2807]^t.$$

S.5: We now go back to the step S.3.

S.3: $\mathbf{c}' = (I - H)\mathbf{c} = [-0.2632\ 0.3509\ -0.3509\ -0.3509]^t$, $s_k = 2.05$.

$$\mathbf{x} = [x_1\ x_3\ x_4\ x_6]^t = [0.75\ 0\ 1\ 1]^t.$$

S.4: We remove x_3 since it has become zero. Hence we remove the second column vector of A and the second element of $\mathbf{c}$. indexarray $= [0\ 2\ 3\ 0\ 5\ 0\ 7]^t$.

$$\mathbf{d} = A^+\mathbf{b} = [0.75\ 1\ 1]^t.$$

S.5: We now go back to step S.3.

S.3: $H = A^+A$ is the unit matrix of order 3. $\mathbf{c}' = [0\ 0\ 0]^t$. Since $c_i' \leq 0$ for all i, we cannot compute s_k.

S.6: Compute $z = \mathbf{c}^t\mathbf{x} = (0 \times 0.75) + (-0.75 \times 1) + (-0.5 \times 1) = -1.25$.
Output $\mathbf{x} = [0.75\ \ 0\ \ 0\ \ 1\ \ 0\ \ 1\ \ 0]^t$, $z = -1.25$.
We subject the solution (basic) to the optimality test. Here

$$B = \begin{bmatrix} 1 & 0.25 & -1 \\ 0 & 0.5 & -0.5 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{c}_B = \begin{bmatrix} 0 \\ -0.75 \\ -0.5 \end{bmatrix}, \quad \mathbf{p}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{p}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$\mathbf{p}_5 = \begin{bmatrix} -8 \\ -12 \\ 0 \end{bmatrix}, \quad \mathbf{p}_7 = \begin{bmatrix} 9 \\ 3 \\ 0 \end{bmatrix}, \quad c_2 = 0, \quad c_3 = 0, \quad c_5 = 20, \quad c_7 = 6.$$

S.7: Compute $\mathbf{y}^t = \mathbf{c}_B^t B^{-1} = [0\ \ -1.5\ \ -1.25]$.

Note: Since B is nonsingular, $B^+ = B^{-1}$ is already available from step S.3.
S.8: Compute $z_2 - c_2 = \mathbf{y}^t\mathbf{p}_2 - c_2 = -1.5, z_3 - c_3 = \mathbf{y}^t\mathbf{p}_3 - c_3 = -1.25$. Similarly z
$c_5 = -2$ and $z_7 - c_7 = -10.5$.

S.9: Since all the values of $z_j - c_j \leq 0$, the optimal solution is reached and is gi
by $\mathbf{x}_B = [x_1\ \ x_4\ \ x_6]^t = B^{-1}\mathbf{b} = [0.75\ \ 1\ \ 1]^t$ or $\mathbf{x} = [0.75\ \ 0\ \ 0\ \ 1\ \ 0\ \ 1\ \ 0]^t$ and $z_B = \mathbf{c}_B^t\mathbf{x}$
$(0 \times 0.75) + (-0.75 \times 1) + (-0.5 \times 1) = -1.25$.

Example 9 (*Two x_i becoming zero simultaneously*)

$$\begin{aligned} &\min z = 2x_1 + 7x_2 - 2x_3 \ \text{ s.t.} \\ &x_1 + 2x_2 + x_3 \leq 1 \\ &-4x_1 - 2x_2 + 3x_3 \leq 2 \\ &x_1, x_2, x_3 \geq 0 \end{aligned}$$

S.1: Input

$$m = 2,\ n = 5,\ A = \begin{bmatrix} 1 & 2 & 1 & 1 & 0 \\ -4 & -2 & 3 & 0 & 1 \end{bmatrix},\ \mathbf{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix},\ \mathbf{c} = [2\ \ 7\ \ -2\ \ 0$$

S.1a: Indexarray $= [0\ \ 0\ \ 0\ \ 0\ \ 0]^t$.

S.2: $\mathbf{d} = A^+\mathbf{b} = [-0.1946\ \ 0.2270\ \ 0.5243\ \ 0.2162\ \ 0.1027]^t, \mathbf{e} = A\mathbf{d} = [1\ \ 2]^t = \mathbf{b}$.
Hence the equation $A\mathbf{x} = \mathbf{b}$ is consistent.

S.3: $\mathbf{c}' = (I - H)\mathbf{c} = [-2.2378\ \ 2.6108\ \ -1.4703\ \ -1.5135\ \ 0.6811]^t$, $s_k = 0.0870$

$$\mathbf{x} = [x_1\ \ x_2\ \ x_3\ \ x_4\ \ x_5]^t = [0\ \ 0\ \ 0.6522\ \ 0.3478\ \ 0.0435]^t.$$

S.4: We now solve the problem removing (i) x_1 only from the basis as well as (ii) keep
x_1 in the basis and removing x_2 from the basis.

Case (i). We remove x_1 from the basis. Hence the first column vector of A as well as
first element of $\mathbf{c}$ are removed.
Indexarray $= [1\ \ 0\ \ 0\ \ 0\ \ 0]^t, \mathbf{d} = A^+\mathbf{b} = [0.0723\ \ 0.6627\ \ 0.1928\ \ 0.1566]^t$.

S.3: $\mathbf{c}' = (I - H)\mathbf{c} = [0.8313\ \ 0.1205\ \ -1.7831\ \ 1.3012]^t$, $s_k = 0.0870$.

$$\mathbf{x} = [x_2\ \ x_3\ \ x_4\ \ x_5]^t = [0\ \ 0.6522\ \ 0.3478\ \ 0.0435]^t.$$

S.4: We remove x_2 since it has become zero. Hence we remove the first column vector of A and the first element of **c**. Indexarray $= [1\ 2\ 0\ 0\ 0]^t$.

$$\mathbf{d} = A^+\mathbf{b} = [0.6364\ \ 0.3636\ \ 0.0909]^t.$$

S.5: We now go back to step S3,

S.3: $\mathbf{c}' = (I - H)\mathbf{c} = [-0.1818\ \ 0.1818\ \ 0.5455]^t$, $s_k = 0.1667$.

$$\mathbf{x} = [x_3\ x_4\ x_5]^t = [0.6667\ \ 0.3333\ \ 0]^t.$$

S.4: We remove x_5 since it has become zero. Hence we remove the last column vector of A and the last element of **c**. indexarray $= [1\ 2\ 0\ 0\ 5]^t$.

$$\mathbf{d} = A^+\mathbf{b} = [0.6667\ \ 0.3333]^t.$$

S.5: We now go back to step S3.

S.3: $H = A^+A$ is the unit matrix of order 2. $\mathbf{c}' = (I - H)\mathbf{c} = [0\ \ 0]^t$. Since $c_i' \le 0$ for all i, we cannot compute s_k.

S.6: Compute $z = \mathbf{c}^t\mathbf{x} = (-2 \times 0.6667) + (0 \times 0.3333) = -1.3333$.
Output $\mathbf{x} = [x_1\ x_2\ x_3\ x_4\ x_5]^t = [0\ \ 0\ \ 0.6667\ \ 0.3333\ \ 0]^t$, $z = -1.3333$.
We subject the solution (basic) to the optimality test. Here

$$B = \begin{bmatrix} 1 & 1 \\ 3 & 0 \end{bmatrix}, \quad \mathbf{p}_1 = \begin{bmatrix} 1 \\ -4 \end{bmatrix}, \quad \mathbf{p}_2 = \begin{bmatrix} 2 \\ -2 \end{bmatrix}, \quad \mathbf{p}_5 = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$
$$\mathbf{c}_B^t = [-2\ \ 0], \quad c_1 = 2, \quad c_2 = 7, \quad c_5 = 0$$

S.7: Compute $\mathbf{y}^t = \mathbf{c}_B^t B^{-1} = [0\ \ 0.6667]$.

Note: Since B is nonsingular, $B^+ = B^{-1}$ is already available from step S.3.

Compute $z_1 - c_1 = \mathbf{y}^t\mathbf{p}_1 - c_1 = [0\ \ -0.6667]\begin{bmatrix} 1 \\ -4 \end{bmatrix} - 2 = 0.6667$.

Similarly, $z_2 - c_2 = -5.6667$ and $z_5 - c_5 = -0.6667$. Since $z_1 - c_1 \ge 0$, the solution is not optimal.

Case (ii). Unlike case (i), here we remove x_2 keeping x_1 in the basis.

Indexarray $= [0\ 2\ 0\ 0\ 0]^t, \mathbf{d} = A^+\mathbf{b} = [0\ \ 0.6364\ \ 0.3636\ \ 0.0909]^t$.

S.5: We now go back to step S.3.

S.3: $\mathbf{c}' = (I - H)\mathbf{c} = [0.0000\ \ -0.1818\ \ 0.1818\ \ 0.5455]^t$, $s_k = 0.1667$.

$$\mathbf{x} = [x_1\ x_3\ x_4\ x_5]^t = [0\ \ 0.6667\ \ 0.3333\ \ 0]^t.$$

Again we are confronted to a situation where there are two zeros in **x**. We apply the same procedure as above, i.e., (a) keep x_5 in the basis and remove x_1 and (b) keep x_1 in the basis and remove x_5.

Subcase (a): We remove x_1 from the basis. Hence the first column vector of A as well as the first element of **c** are removed. indexarray $= [1\ 2\ 0\ 0\ 0]^t$.

$$\mathbf{d} = A^+\mathbf{b} = [0.6364\ \ 0.3636\ \ 0.0909]^t.$$

S.5: We now go back to the step S3.

S.3: $\mathbf{c}' - (I - H)\mathbf{c} = [-0.1818\ \ 0.1818\ \ 0.5455]^t$, $s_k = 0.1667$.

$$\mathbf{x} = [x_3\ \ x_4\ \ x_5]^t = [0.6667\ \ 0.3333\ \ 0]^t.$$

S.4: We remove x_5 since it has become zero. Hence we remove the last column vector of A and the last element of **c**. indexarray $= [1\ \ 2\ \ 0\ \ 0\ \ 5]^t$.

After this step, the calculation proceeds in the same way as in case (i).

Subcase (b): Unlike subcase (a), here we remove x_5, keeping x_1 in the basis. Indexarray $= [0\ \ 2\ \ 0\ \ 0\ \ 5]^t, \mathbf{d} = A^+\mathbf{b} = [-0.0135\ \ 0.6486\ \ 0.3649]^t$.

S.5: We now go back to the step S3.

S.3: $\mathbf{c}' = (I - H)\mathbf{c} = [-0.0811\ \ -0.1081\ \ 0.1892]^t$, $s_k = 1.9286$.

$$\mathbf{x} = [x_1\ \ x_3\ \ x_4]^t = [0.1429\ \ 0.8571\ \ 0]^t.$$

S.4: We remove x_4 since it has become zero. Hence we remove the last column vector of A and the last element of **c**. indexarray $= [0\ \ 2\ \ 0\ \ 4\ \ 5]^t$.

$$\mathbf{d} = A^+\mathbf{b} = [0.1429\ \ 0.8571]^t.$$

S.5: We now go back to the step S3.

S.3: $H = A^+A$ is the unit matrix of order 2. $\mathbf{c}' = (I - H)\mathbf{c} = [0\ \ 0]^t$. Since $c'_i \le 0$ for all i, we cannot compute s_k.

S.6: Compute $z = \mathbf{c}^t\mathbf{x} = (2 \times 0.1429) + (-2 \times 0.8571) = -1.4286$.

Output $\mathbf{x} = [x_1\ \ x_2\ \ x_3\ \ x_4\ \ x_5]^t = [0.1429\ \ 0\ \ 0.8571\ \ 0\ \ 0]^t$, $z = -1.4286$.

We subject the solution (basic) to the optimality test. Here

$$B = \begin{bmatrix} 1 & 1 \\ -4 & 3 \end{bmatrix}, \quad \mathbf{p}_2 = \begin{bmatrix} 2 \\ -2 \end{bmatrix}, \quad \mathbf{p}_4 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \mathbf{p}_5 = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

$$\mathbf{c}_B^t = [2\ \ -2], \quad c_2 = 7, \quad c_4 = 0, \quad c_5 = 0.$$

Note: Since B is nonsingular, $B^+ = B^{-1}$ is already available from step S.3.

S.7: Compute $\mathbf{y}^t = \mathbf{c}_B^t B^{-1} = [-0.2857\ \ -0.5714]$.

S.8: Compute $z_2 - c_2 = \mathbf{y}^t\mathbf{p}_2 - c_2 = -6.4286$, $z_4 - c_4 = \mathbf{y}^t\mathbf{p}_4 - c_4 = -0.2857$ and $z_5 - c_5 = \mathbf{y}^t\mathbf{p}_5 - c_5 = -0.5714$. Since all the values of $z_j - c_j \le 0$, the optimal solution is reached. Therefore, the required optimal solution is $\mathbf{x} = [x_1\ \ x_2\ \ x_3\ \ x_4\ \ x_5]^t = [0.1429\ \ 0\ \ 0.8571\ \ 0\ \ 0]^t$ and the objective function value is $z = -1.4286$.

Example 10 (*Optimal solution not reached – DHALP failed*)

$$\begin{aligned} &\min z = -3x_1 - 2x_2 \ \text{s.t.} \\ &2x_1 + x_2 \le 2 \\ &3x_1 + 4x_2 \ge 2 \\ &x_1, x_2 \ge 0. \end{aligned}$$

S.1: Input $m = 2, n = 4, A = \begin{bmatrix} 2 & 1 & 1 & 0 \\ 3 & 4 & 0 & -1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$, $\mathbf{c} = [-3\ \ -2\ \ 0\ \ 0]^t$.

S.1a: Indexarray $= [0\ \ 0\ \ 0\ \ 0]^t$.

S.2: $\mathbf{d} = A^+\mathbf{b} = [0.7143\ \ 0.0000\ \ 0.5714\ \ 0.1429]^t, \mathbf{e} = A\mathbf{d} = [2\ \ 2]^t = \mathbf{b}$. Hence the equation $A\mathbf{x} = \mathbf{b}$ is consistent.

S.3: $\mathbf{c}' = (I - H)\mathbf{c} = [-0.4643\ \ 0.2500\ \ 0.6786\ \ -0.3929]^t$, $s_k = 0$.

$$\mathbf{x} = [x_1\ \ x_2\ \ x_3\ \ x_4]^t = [0.7143\ \ 0\ \ 0.5714\ \ 0.1429]^t.$$

S.4: We remove x_2 since it has become zero. Hence we remove the second column vector of A and the second element of $\mathbf{c}$. indexarray $= [0\ \ 2\ \ 0\ \ 0]^t$.

$$\mathbf{d} = A^+\mathbf{b} = [0.7143\ \ 0.5714\ \ 0.1429]^t.$$

S.5: We now go back to step S.3.

S.3: $\mathbf{c}' = (I - H)\mathbf{c} = [-0.2143\ \ 0.4286\ \ -0.6429]^t$, $s_k = 1.3333$.

$$\mathbf{x} = [x_1\ \ x_3\ \ x_4]^t = [1.0000\ \ 0.0000\ \ 1.0000]^t.$$

S.4: We remove x_3 since it has become zero. Hence we remove the second column vector of A and the second element of $\mathbf{c}$. indexarray $= [0\ \ 2\ \ 3\ \ 0]^t$.

$$\mathbf{d} = A^+\mathbf{b} = [1\ \ 1]^t.$$

S.5: We now go back to step S.3.

S.3: $H = A^+A$ is the unit matrix of order 2. $\mathbf{c}' = (I - H)\mathbf{c} = [0\ \ 0]^t$. Since $c_i \le 0$ for all i, we cannot compute s_k.

S.6: Compute $z = \mathbf{c}^t\mathbf{x} = (-3 \times 1) + (0 \times 1) = -3$.
Output $\mathbf{x} = [x_1\ \ x_2\ \ x_3\ \ x_4]^t = [1\ \ 0\ \ 0\ \ 1]^t$, $z = -3$.
We subject the solution (basic) to the optimality test. Here

$$B = \begin{bmatrix} 2 & 0 \\ 3 & -1 \end{bmatrix},\ \mathbf{p}_2 = \begin{bmatrix} 1 \\ 4 \end{bmatrix},\ \mathbf{p}_3 = \begin{bmatrix} 1 \\ 0 \end{bmatrix},\ \mathbf{c}_B^t = [-3\ \ \ 0],\ c_2 = -2,\ c_3 = 0.$$

S.7: Compute $\mathbf{y}^t = \mathbf{c}_B^t B^{-1} = [-1.5\ \ 0]$.

S.8: Compute $z_2 - c_2 = \mathbf{y}^t\mathbf{p}_2 - c_2 = 0.5$, $z_3 - c_3 = \mathbf{y}^t\mathbf{p}_3 - c_3 = -1.5$.
Since $z_2 - c_2 > 0$, the solution is not optimal: either the solution is unbounded or the DHALP could not reach the optimal solution. We use the revised SA to proceed further, starting with the output solution $\mathbf{x}$ of the DHALP algorithm as the initial basic feasible solution.

Revised S.4: From the DHALP, the nonbasic variables are x_2 and x_3.

Step A. Determining the entering vector, $\mathbf{p}_j$
Compute $z_j - c_j$ for nonbasic $\mathbf{p}_2$ and $\mathbf{p}_3$. From the step S.8, $z_2 - c_2 = 0.5$ and $z_3 - c_3 = -1.5$. Since $z_2 - c_2$ is positive, x_2 now becomes a basic variable and $\mathbf{p}_2$ enters the basis.

Step B. Determining the leaving vector $\mathbf{p}_j$
The current basic solution is

(i) $$\mathbf{x}_B = \begin{bmatrix} x_1 \\ x_4 \end{bmatrix} = B^{-1}\mathbf{b} = \begin{bmatrix} 0.5 & 0 \\ 1.5 & -1 \end{bmatrix}\begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

(ii) The constraint coefficients of the entering variables, i.e.,

$$\alpha^{(2)} = B^{-1}\mathbf{p}_2 = [0.5 - 2.5]^t.$$

$\theta = \min\{1/0.5, -\} = 2$. Thus x_1 leaves the basis. We go back to the step A.

Step A.

$$B = \begin{bmatrix} 1 & 0 \\ 4 & -1 \end{bmatrix}, \quad \mathbf{p}_1 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \quad \mathbf{p}_3 = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

$$\mathbf{c}_B^t = [-2\ \ 0], \quad c_1 = -3, \quad c_3 = 0.$$

Compute $\mathbf{y}^t = \mathbf{c}_B^t B^{-1} = [-2\ \ 0]$. Compute $z_j - c_j$ for nonbasic $\mathbf{p}_1$ and $\mathbf{p}_3$.

$$z_1 - c_1 = \mathbf{y}^t\mathbf{p}_1 - c_1 = -1 \text{ and } z_3 - c_3 = -2.$$

Since all the values of $z_j - c_j \leq 0$, the optimal solution has been reached and is give by $\mathbf{x}_B = [x_2\ \ x_4]^t = B^{-1}\mathbf{b} = [2\ \ 6]^t$ or $\mathbf{x} = [x_1\ \ x_2\ \ x_3\ \ x_4]^t = [0\ \ 2\ \ 0\ \ 6]^t$ and $z_B = \mathbf{c}_B^t\mathbf{x}_B =$ $(-2 \times 2) + (0 \times 6) = -4$.

4. Conclusions

Versus simplex algorithm: The direct heuristic algorithm (DHALP) and the most popula and most widely used simplex method (SA) have the following differences.

If the SA does not enter into an infinite loop [1, 3, 40], i.e. cycling – a situation mostl not encountered in practice – then it will certainly provide the required solution of th LPP. The DHALP may not certainly provide the solution although in most (over 95% o the problems solved by us) problems it does. Even if it does not provide the optima solution, it will usually provide one close to the optimal one. One may obtain the optima one from this solution using the revised SA in a fewer steps.

The bottom-most row, viz., the $-c_j$ row in the SA tells us whether the optimal solutio is reached or not while, in the DHALP, the optimality test based on using the coefficient of nonbasic variables checks the optimality.

Each next-tableau of the SA is computed by the elementary row/column operation similar to those in Gauss reduction method for solving linear systems. Each tableau need almost the same amount of computation. The coefficient matrix A in the DHALP loose in each step one column corresponding to the variable x_i whose value becomes zer resulting in successive reduction in computations. In addition, a minimum norm leas squares inverse (p-inverse) of the shrunk matrix A is calculated from the current A^+ an the current A in $O(mn)$ operations and not from the current A alone in $O(mn^2)$ operations

A variable in the SA may enter into the basis and may go out; this may happen number of times. In the DHALP, once a variable that leaves the basis will never enter int the basis.

The SA is exponential time (in the worst case) while the DHALP is polynomial-tim direct and needs $O(mn^2)$ operations like the algorithms for solving linear systems usin for example, the Gauss reduction method.

While the SA is iterative, and the precise amount of computation in it is not known *priori*, the DHALP is noniterative and the precise amount of computation in it is almos always known beforehand.

Artificial variables for 'equal to' and 'greater than or equal to' constraints are neede in the SA (besides surplus/slack variables) for consistency check. No artificial variable i needed in the DHALP – consistency check is in-built.

Initial basic feasible solution is known in SA while it is unknown in the DHAL *Obtaining a (basic feasible) nonnegative solution of linear systems*: If the DHALP fails t provide an optimal solution for an LPP, it will almost always provide at least nonnegative solution, i.e., a point inside the polytope defined by $A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}$ – which i a very good basic feasible solution from which optimal solution could be obtained within a few iterations of the revised SA.

Versus interior-point methods: Several interior-point methods [13, 12, 36, 19] have been proposed during the last two decades for the LPP. All these methods find a point (a basic feasible solution) inside the polytope $A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}$ and then proceed in the direction of the $\mathbf{c}$-vector in search of the corner (of the polytope) that represents the optimal solution. In the SA, however, we go along a hyperplane to a corner until an optimal corner is found. All these interior-point methods and the SA are iterative. The DHALP attempts to go inside the polytope $A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}$ and then tries to hit upon the optimal solution mathematically noniteratively.

Versus other algorithms: The inequality sorting algorithm [15, 17] and the Barnes algorithm for detection of basic variables [2, 33] are distinctly different from the DHALP in that these are iterative and deterministic.

Error-free computation with DHALP. The DHALP involves only the basic arithmetic operations, viz., add, subtract, multiply, and divide operations. It does not need square-rooting or other operations that produce irrational numbers. A bound [9, 37–39] on the solution vector $\mathbf{x}$ for the LPP can be easily deduced. So the error-free computation using multiple-modulus residue arithmetic or p-adic arithmetic can be implemented on the DHALP. This implementation will appear elsewhere. Most interior-point methods are not amenable to error-free computations as they involve square-rooting operations.

Degenerate, unbounded, infeasible LPPs: We have considered such LPPs in the numerical examples (§ 3) and necessary discussions are included there. However, the DHALP has performed well in all these LPPs.

Small problems: We have attempted to solve small problems in which the constraint matrix A has order not greater than 20. This is mainly because of the limitation of the available computing resources including the limitation of the accuracy of computation in a PC-MATLAB at our disposal. However, a double- or multiple-precision high-level language may be used to solve reasonably large problems with a fair amount of accuracy using the DHALP. We will attempt encoding the DHALP into a high-level program (which is a significant computational problem) for large sparse LPPs in a future article.

Open problem: The problem of finding the necessary and sufficient condition under which the DHALP will produce an optimal solution of the LPP (or, equivalently, the DHALP becomes a direct algorithm) is open. A visualization of the problem geometrically may, however, provide a meaningful sufficient condition without much difficulty. Such a condition could be of some practical use but the necessary and sufficient condition, if found, will be of significant practical importance.

Acknowledgment

The authors wish to thank the authority of the University of Mauritius for providing necessary computing and office facilities for this research while the second author thanks the Tertiary Education Commission for a scholarship.

References

[1] Bazaraa M S, Jarvis J J and Sheraldi H D, Linear Programming and Network Flows, 2nd ed. (Singapore: John Wiley and sons, Inc.) (1990) pp. 165–167

[2] Barnes E R, A variation of Karmarkar's algorithm for solving linear programming problems, *Math. Prog.* **36** (1986) 174–182

[3] Beale E M L, Cycling in the dual simplex algorithm. *Naval Research Logistics Quarterly* (1955) 269–275

[4] Ben Israel A and Greville T N E, Generalized Inverses: Theory and Applications (New Yor Wiley) (1974)

[5] Campbell S L, On the continuity of the Moore-Penrose and Drazkin generalized inverse *Linear algebra and its applications* **18** (1977) 53–57

[6] Dantzig G, Linear Programming and Extensions (New Jersey: Princeton University Pres Princeton) (1963)

[7] Dongarra J J, Du Croz J J, Hammerling S J and Hanson R J, An extended set of Fortran bas linear algebra subprograms, *ACM Trans. Math. Software* **14(1)** (1988) 1–17

[8] Golub G and Kahan W, Calculating the singular values and the pseudoinverse of a matri *SIAM J. Numer. Anal.* **B-2** (1965) 205–224

[9] Gregory R T and Krishnamurthy E V, Methods and Applications of Error-free Computati (New York: Springer Verlag) (1984)

[10] Greville T N E, The pseudoinverse of a rectangular or singular matrix and its application the solution of systems of linear equations, *SIAM Rev.* **1** (1959) 38–43

[11] Hooker J N, Karmarkar's Linear Programming Algorithm. *Interfaces* **16(4)** (1986) 75–90

[12] Karmarkar N, A new polynomial-time algorithm for linear programming, *Tech. Rep* (Ne Jersey: AT&T Bell Labs.) (1984)

[13] Khachiyan L G, A polynomial algorithm in linear programming, *Doklady Akad. Nauk SS* **S244** (1979) 1093–1096

[14] Krishnamurthy E V and Sen S K, Numerical Algorithms: Computations in Science a Engineering (New Delhi: Affiliated East-West Press) (1993)

[15] Lakshmikantham V, Sen S K and Sivasundaram S, An inequality sorting algorithm for a cla of linear programming problems, *J. Math. Anal. Appl.* **174** (1993) 450–460

[16] Lakshmikantham V, Sen S K and Howell G, Vectors versus matrices: p-inversic cryptographic application, and vector implementation, *Neural, Parallel, and Scienti Computations* **4** (1996) 129–140

[17] Lakshmikantham V, Maulloo A K, Sen S K and Sivasundaram S, Solving Line Programming Problems Exactly, *Appl. Maths. Comput.* **81** (1997) 69–87

[18] Lord E A, Sen S K and Venkaiah V Ch, A concise algorithm to solve under over determin linear systems, *Simulation* **54** (1990) 239–240

[19] Lord E A, Venkaiah V Ch and Sen S K, A shrinking polytope method for linear programmi *Neural, Parallel and Scientific Computations* **4** (1996) 325–340

[20] Natick M A, *MATLAB User's Guide* (The Math Works Inc.) (1992)

[21] Moore E H, On the reciprocal of general algebraic matrix (abs.), *Bull Am. Math. Soc.* (1920) 394–395

[22] Murty K G, Linear Complementarity, Linear and Nonlinear Programming, (Germa Heldermann Verlag, Berlin) (1989)

[23] Nazareth J L, Computer Solutions of Linear Programs (Monographs on Numerical Analys (Oxford: Oxford University Press) (1987) pp. 39–40

[24] Parker G and Rardin R, Discrete Optimization, (San Diego: Academic Press) (1988)

[25] Penrose R, A generalized inverse for matrices, *Proc. Chamb. Phil. Soc.* **51** (1955) 406–4

[26] Penrose R, On best approximate solutions of linear matrix equations, *Proc. Chamb. Phil. S* **52** (1956) 17–19

[27] Peters G and Wilkinson J H, The least squares problem and pseudo-inverses, *The Compute* **13** (1970) 309–316

[28] Rao C R and Mitra S K, Generalized Inverse of Matrices and Its Applications (New Yo Wiley) (1971)

[29] Renegar J, A polynomial-time algorithm, based on Newton's method for linear programmi *Math. Prog.* **40** (1988) 59–93

[30] Sen S K and Krishnamurthy E V, Rank-augmented LU-algorithm for computing generaliz matrix inverses, *IEEE Trans. Comput.* **C-23** (1974) 199–201

[31] Sen S K and Prabhu S S, Optimal iterative schemes for computing Moore-Penrose mat inverse, *Int. J. Systems Sci.* **8** (1976) 748–753

[32] Sen S K, Hongwei Du and Fausett D W, A center of a polytope: an expository review an parallel implementation. *Internat. J. Math. Math. Sci.* **16** (2), (1993) 209–224

[33] Sen S K, Sivasundaram S and Venkaiah Ch, Barnes algorithm for linear programming: on detection of basic variables. *Nonlinear World,* **2** (1995)

[34] Stewart G W, On the continuity of the generalized inverse, *SIAM J. Appl. Math.* **17** (1969) 35–45

[35] Taha H A, Operations Research – An Introduction, 4th ed. (New York: Macmilan Publishing Company) (1989) 254–259

[36] Vaidya P M, An algorithm for linear programming which requires $O(((m+n)n^2+(m+n)^{1.5}n)L)$ arithmetic operations, *Proc. ACM Annual Symposium on Theory of Computing* (1987) pp 29–38; also *Math. Prog.* **47** (1990) 175–201

[37] Venkaiah V Ch and Sen S K, A floating-point-like modular arithmetic for polynomials with application to rational matrix processors, *Advances in Modelling and Simulation* **9(1)** (1987) 1–12

[38] Venkaiah V Ch and Sen S K, Computing a matrix symmetrizer exactly using modified multiple modulus residue arithmetic, *J. Comput. Appl. Math.* **21** (1988) 27–40

[39] Venkaiah V Ch and Sen S K, Error-free matrix symmetrizers and equivalent symmetric matrices, *Acta. Appl. Math.* **21** (1990) 291–313

[40] Wagner H M, Principles of Operations Research 2nd ed. (USA: Prentice Hall, Inc.) (1969) pp. 115–116

Proc. Indian Acad. Sci. (Math. Sci.), Vol. 110, No. 1, February 2000, pp. 103–115.
© Printed in India

An approximate solution for spherical and cylindrical piston problem

S K SINGH and V P SINGH

Centre for Aeronautical System Studies and Analyses, New Thippasandra P. O.,
Bangalore 560 075, India
E-mail. drvpsingh@hotmail.com

MS received 28 December 1998; revised 10 September 1999

Abstract. A new theory of shock dynamics (NTSD) has been derived in the form of a finite number of compatibility conditions along shock rays. It has been used to study the growth and decay of shock strengths for spherical and cylindrical pistons starting from a non-zero velocity. Further a weak shock theory has been derived using a simple perturbation method which admits an exact solution and also agrees with the classical decay laws for weak spherical and cylindrical shocks.

Keywords. Shock dynamics; compatibility conditions; blast wave; weak shock propagation.

1. Introduction

Though the occurrence of a shock discontinuity in compressible flows and the jump conditions across it are known for more than a century, the idea of deriving an infinite system of compatibility conditions along shock rays (Prasad [9]) was discovered only recently (Grinfeld [2], Maslov [7]). By truncating the infinite system of compatibility conditions at an appropriate level, a new theory of shock dynamics (NTSD) has been proposed (Ravindran and Prasad [11]) which enables one to compute the position and strength of a shock front and also to determine the flow behind the shock up to a short distance. Lazarev, Prasad and Singh [6] used NTSD to study the growth and decay of a plane shock originating due to an accelerating or decelerating piston. They also compared the results from NTSD with those from Harten's total variation diminishing (TVD) finite difference scheme (FDM) and found good agreement. It was also noted that NTSD consumes only 0.5% of the computational time taken by FDM, while giving almost the same (and in some cases even better) accuracy for the solution.

The problem of blast wave propagation originating from the detonation of an explosive has been modeled as that of a symmetrically expanding spherical or cylindrical piston by several authors (see Stanyukovich [15], Courant and Friedrichs [1]). This problem presents an example of a flow field in which the flow behind the shock front is highly non-uniform due to a rapid decay of the flow behind the shock, which makes the use of Whitham's shock dynamics [18, 19] (which ignores the effects of the flow behind the shock) inapplicable for such problems.

Another conventional approach for solving blast problem has been either the use of self-similar solutions, valid only for a short time and a short distance from the site of explosion (Taylor [16], Sedov [12]) or resorting to specially devised finite difference schemes such as those due to Glimm or Godunov (see Holt [3], Peyret and Taylor [8]).

The relative efficiency of NTSD over a self-similar solution and a finite difference solution has been examined by Singh and Singh [13] in the case of a single conservation law.

In this paper, we apply NTSD to obtain an approximate solution to spherical and cylindrical piston problem. A general approach for both accelerating and decelerating cases has been presented, in particular the latter can model closely the phenomenon of rapid decay of the flow behind the shock occurring in a blast. Further, a theory for the propagation of weak shocks has been derived by using a simple perturbation of the ordinary differential equations appearing in the NTSD. This solution reduces to the well known classical results for the decay of weak spherical and cylindrical shocks.

2. Dynamical compatibility conditions

The unsteady flow of an ideal gas with constant specific heats for spherical or cylindrical symmetry is given by the following system

$$\begin{aligned}
&\frac{\partial \rho}{\partial t}+u\frac{\partial \rho}{\partial r}+\rho\frac{\partial u}{\partial r}+j\frac{\rho u}{r}=0,\\
&\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial r}+\frac{1}{\rho}\frac{\partial p}{\partial r}=0,\\
&\frac{\partial p}{\partial t}+u\frac{\partial p}{\partial r}+\gamma p\frac{\partial u}{\partial r}+j\frac{\gamma p u}{r}=0,
\end{aligned}\tag{2.1}$$

where ρ, u, p are the density, velocity and the pressure of the gas, γ is the ratio of specific heats; t, r are the time and radial co-ordinates respectively and $j = 1, 2$ for cylindrical and spherical cases respectively.

Let $r = R(t)$ be the position of the shock front propagating into the gas at uniform state and at rest ahead of the shock front. We introduce the following notations

$$D_0 = [\rho], \quad H_0 = [u], \quad S_0 = [p], \tag{2.2}$$

where $[\,]$ denotes the jump across the shock front: $[G] = G_+ - G_-$, $+$ denotes the state ahead and $-$ that behind the shock front. The expressions for H_0, S_0 and the shock velocity C are given by the well known Rankine–Hugoniot relations

$$\begin{aligned}
&H_0 = CD_0(\rho_+ - D_0)^{-1}, \quad S_0 = \rho_+ D_0 C^2(\rho_+ - D_0)^{-1},\\
&C = a_+(2(\rho_+ - D_0)(2\rho_+ + (\gamma-1)D_0)^{-1})^{1/2},
\end{aligned}\tag{2.3}$$

where a_+ is the local sound velocity ahead of the shock. It is assumed that the flow variables $\rho(r,t), u(r,t)$ and $p(r,t) \in C^\infty$ behind the shock front. The following relations can be written for the derivatives of the flow variables on the shock front

$$\frac{\mathrm{d}Z}{\mathrm{d}t}=\frac{\partial Z}{\partial t}+C\frac{\partial Z}{\partial r}, \quad \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial^N Z}{\partial r^N}\right)=\frac{\partial^{N+1}Z}{\partial t\partial r^N}+C\frac{\partial^{N+1}Z}{\partial r^{N+1}} \tag{2.4}$$

for $N = 1, 2, 3, \ldots$, and $Z(r,t)$ can be any of the flow variables ρ, u or p. Taking jump on both sides in (2.4), we obtain

$$\left[\frac{\partial Z}{\partial t}\right]=\frac{\mathrm{d}}{\mathrm{d}t}[Z]-C\left[\frac{\partial Z}{\partial r}\right], \quad \left[\frac{\partial^{N+1}Z}{\partial t\partial r^N}\right]=\frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{\partial^N Z}{\partial r^N}\right]-C\left[\frac{\partial^{N+1}Z}{\partial r^{N+1}}\right]. \tag{2.5}$$

We notice that the first equation in (2.5) is the first kinematical compatibility condition of Thomas [17]. Replacing Z by ρ, u, p in (2.5), we get the following relations

$$\begin{aligned}
\left[\frac{\partial\rho}{\partial t}\right] &= \frac{\mathrm{d}}{\mathrm{d}t}D_0 - CD_1, & \left[\frac{\partial^{N+1}\rho}{\partial t\partial r^N}\right] &= \frac{\mathrm{d}}{\mathrm{d}t}D_N - CD_{N+1},\\
\left[\frac{\partial u}{\partial t}\right] &= \frac{\mathrm{d}}{\mathrm{d}t}H_0 - CH_1, & \left[\frac{\partial^{N+1}u}{\partial t\partial r^N}\right] &= \frac{\mathrm{d}}{\mathrm{d}t}H_N - CH_{N+1},\\
\left[\frac{\partial p}{\partial t}\right] &= \frac{\mathrm{d}}{\mathrm{d}t}S_0 - CS_1, & \left[\frac{\partial^{N+1}p}{\partial t\partial r^N}\right] &= \frac{\mathrm{d}}{\mathrm{d}t}S_N - CS_{N+1},
\end{aligned} \tag{2.6}$$

where

$$D_N = \left[\frac{\partial^N\rho}{\partial r^N}\right], \quad H_N = \left[\frac{\partial^N u}{\partial r^N}\right], \quad S_N = \left[\frac{\partial^N p}{\partial r^N}\right]. \tag{2.7}$$

We consider the jump of the left hand side of (2.1) across the shock front. Using (2.6), we obtain the first set of dynamic compatibility conditions

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{U}_0 + \mathbf{P}\cdot\mathbf{U}_1 = \mathbf{f}_0, \tag{2.8}$$

where

$$\mathbf{U}_N = \begin{pmatrix} D_N \\ H_N \\ S_N \end{pmatrix}, \quad N = 1, 2, \ldots \tag{2.9}$$

and

$$\mathbf{P} = \begin{pmatrix} -(C+H_0) & \rho_+ - D_0 & 0 \\ 0 & -(C+H_0) & (\rho_+ - D_0)^{-1} \\ 0 & \gamma(p_+ - S_0) & -(C+H_0) \end{pmatrix} \tag{2.10}$$

and

$$\mathbf{f}_0 = -jr^{-1}\begin{pmatrix} H_0(\rho_+ - D_0) \\ 0 \\ \gamma H_0(p_+ - S_0) \end{pmatrix}. \tag{2.11}$$

To derive the second set of compatibility conditions, we differentiate (2.1) with respect to r and take the jump of the left hand side of the resulting equation. Using (2.6), we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{U}_1 + \mathbf{P}\cdot\mathbf{U}_2 = \mathbf{f}_1, \tag{2.12}$$

where

$$\mathbf{f}_1 = \begin{pmatrix} 2D_1H_1 + jr^{-1}(H_0D_1 - (\rho_+ - D_0)H_1) + jr^{-2}(\rho_+ - D_0)H_0 \\ H_1^2 - S_1D_1(\rho_+ - D_0)^{-2} \\ (\gamma+1)H_1S_1 + jr^{-1}\gamma(H_0S_1 - (p_+ - S_1)H_1) + jr^{-2}\gamma(p_+ - S_0)H_0 \end{pmatrix}. \tag{2.13}$$

(It may be noted that for $j = 0$, the system of compatibility conditions (2.8) and (2.12) reduces to that for plane shocks, see [6] for details.)

Repeating the same procedure, an infinite system of compatibility conditions can be derived in the following general form

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{U}_N + \mathbf{P}\cdot\mathbf{U}_{N+1} = \mathbf{f}_N, \quad N = 0, 1, 2\ldots \tag{2.14}$$

The equations (2.8) and (2.11) are the first two members of the system (2.14).

It is obvious that for computational purposes, it is more convenient to work with a scalar system of equations. We now describe a procedure to reduce the above system of vector compatibility conditions into an equivalent system of scalar compatibility conditions. We note that the eigenvalues of the matrix **P** are

$$\lambda_1 = -(C + H_0), \quad \lambda_{2,3} = -(C + H_0) \pm a_-, \tag{2.15}$$

where a_- is the sound velocity behind the shock front. As the shock strength H_0 tends to zero, the shock velocity and the local sound velocity tend to a common value, say a_0, hence λ_2 tends to zero. In this limit the first set of compatibility conditions (2.8) must lead to the characteristic compatibility condition in which the derivative terms must be zero. Hence, we choose the left eigenvector **L** corresponding to λ_2:

$$\mathbf{L} = (0, (\rho_+ - D_0)/2, (2a_-)^{-1})$$

and introduce the following notations

$$\begin{aligned} \pi_0 &= D_0, \\ \pi_N &= \mathbf{L}\cdot\mathbf{U}_N = (\rho_+ - D_0)H_N/2 + S_N(2a_-)^{-1}, \quad N = 1, 2, \ldots \end{aligned} \tag{2.16}$$

Multiplying (2.8) by **L**, we get after some simplifications

$$\frac{\mathrm{d}\pi_0}{\mathrm{d}t} = -g\left(\lambda_2\pi_1 + \frac{j}{2r}Ca\pi_0\right), \tag{2.17}$$

where

$$g = ((\rho_+ - \pi_0)a/2 + b(2a_-)^{-1})^{-1}, \tag{2.18}$$

$$a = \frac{\partial H_0}{\partial \pi_0} = \frac{C\rho_+(4\rho_+ + (\gamma - 3)\pi_0)}{2(\rho_+ - \pi_0)^2(2\rho_+ + (\gamma - 1)\pi_0)}, \tag{2.19}$$

$$b = \frac{\partial S_0}{\partial \pi_0} = 4\gamma p_+\rho_+(2\rho_+ + (\gamma - 1)\pi_0)^2)^{-2}. \tag{2.20}$$

Next, to express D_1, H_1, S_1 in terms of π_0 and π_1, we multiply (2.8) by $\mathbf{P}^{-1}$ to get

$$\mathbf{U}_1 = \mathbf{P}^{-1}\left(\mathbf{f}_0 - \frac{\mathrm{d}\mathbf{U}_0}{\mathrm{d}t}\right). \tag{2.21}$$

Using (2.17), we get from (2.21)

$$\begin{aligned} D_1 &= jr^{-1}Cd_{10}\pi_0 + gd_{11}\pi_1, \quad H_1 = jr^{-1}Ch_{10}\pi_0 + h_{11}\pi_1, \\ S_1 &= jr^{-1}Cs_{10}\pi_0 + s_{10}\pi_1, \end{aligned} \tag{2.22}$$

where

$$\begin{aligned} d_{10} = {} & \lambda_1^{-1}(-1 + ga_-/2) - aga_-(2\lambda_2\lambda_3)^{-1}(\rho_+ - \pi_0) \\ & + (\lambda_1\lambda_2\lambda_3)^{-1}(-a_-^2 + bga_-/2), \end{aligned}$$

$$\begin{aligned}
d_{11} &= \lambda_2\lambda_1^{-1} - a\lambda_3^{-1}(\rho_+ - \pi_0) + b(\lambda_1\lambda_3)^{-1},\\
h_{10} &= (\lambda_2\lambda_3)^{-1}(aga_-\lambda_1/2 - (-a_-^2 + bga_-/2)(\rho_+ - \pi_0)^{-1}),\\
h_{11} &= \lambda_3^{-1}(\lambda_1 a - b(\rho_+ - \pi_0)^{-1}),\\
s_{10} &= (\lambda_2\lambda_3)^{-1}(-\gamma(p_+ - S_0)aga_-/2 + \lambda_1(-a_-^2 + bga_-/2)),\\
s_{11} &= \lambda_3^{-1}(-\gamma(p_+ - S_0)a + \lambda_1 b).
\end{aligned} \tag{2.23}$$

To reduce the second compatibility condition into scalar form, we multiply the equation (2.12) by **L** to obtain

$$\frac{\mathrm{d}\pi_1}{\mathrm{d}t} = \delta_1\pi_0^2 + \delta_2\pi_1^2 + \delta_3\pi_0\pi_1 + \delta_4\pi_0 + \delta_5\pi_1 - \lambda_2\pi_2, \tag{2.24}$$

where

$$\begin{aligned}
\delta_1 &= \left(\frac{jC}{r}\right)^2 \left\{\frac{1}{2}(\rho_+ - \pi_0)(h_{10}^2 - (\rho_+ - \pi_0)^{-2}s_{10}d_{10}) + \frac{s_{10}}{2a_-}((\gamma + 1)h_{10}\right.\\
&\quad \left. + \gamma(\rho_+ - \pi_0)^{-1}) + \frac{1}{4}ga_-(h_{10} - es_{10})\right\},\\
\delta_2 &= \frac{1}{2}g^2\left\{(\rho_+ - \pi_0)(h_{11}^2 - (\rho_+ - \pi_0)^{-2}) + (\gamma+1)a_-^{-1}h_{11}s_{11} + \lambda_2(h_{11} - es_{11})\right\},\\
\delta_3 &= \frac{j}{2r}Cg\left\{(\rho_+ - \pi_0)(2h_{10}h_{11} - (\rho_+ - \pi_0)^{-2}(s_{10}d_{11} + s_{11}d_{10}) + (\gamma + 1)a_-^{-1}\right.\\
&\quad \left. \times(h_{10}s_{11} + h_{11}s_{10}) + \gamma s_{11}a_-^{-1}(\rho_+ - \pi_0)^{-1} + \frac{2}{g}\lambda_2(h_{10} - es_{10}) + a_-(h_{11} - es_{11})\right\},\\
\delta_4 &= \frac{j}{2r^2}Ca_-(1 - jh_{10}(\rho_+ - \pi_0)), \quad \delta_5 = -\frac{j}{2r}ga_-h_{11}(\rho_+ - \pi_0)
\end{aligned} \tag{2.25}$$

and

$$e = -\frac{1}{a_-^2}\frac{\partial a_-}{\partial \pi_0} = (2a_-(\rho_+ - \pi_0))^{-1}\left(\frac{\gamma\rho_+C^4}{a_-^2a_+^2(\rho_+ - \pi_0)^2} - 1\right). \tag{2.26}$$

We add the equation of the shock path to the above system of compatibility conditions

$$\frac{\mathrm{d}R}{\mathrm{d}t} = C, \tag{2.27}$$

where C is the shock velocity given by (2.3).

The new theory of shock dynamics (NTSD) using compatibility conditions up to the second order is obtained by putting $\pi_2 = 0$ in (2.24). The NTSD is valid for a shock of arbitrary strength.

3. Initial conditions for accelerating or decelerating piston problem

In this section we derive the initial conditions to solve the set of ordinary differential equations (2.17) and (2.24) (with π_2 set equal to zero) to obtain an approximate solution for an expanding spherical or cylindrical piston. We consider the flow produced by a spherical or cylindrical piston expanding with a non-zero positive velocity and a non-zero positive or negative acceleration into a gas at rest ahead of the piston. Let the piston position at time t be $R_p(t)$. Mathematically, the problem consists in solving the system of equations (2.1) with the following initial and boundary conditions

$$u(r,0) = 0, \quad p(r,0) = p_+, \quad \rho(r,0) = \rho_+ \quad \text{for } r > 0 \tag{3.1}$$

and

$$u(R_p(t),t) = R'_p(t) \ (= \text{piston velocity}), \quad \text{for } t > 0. \tag{3.2}$$

The flow variables are non-dimensionalized as follows

$$r = r_0\bar{r}, \quad t = \frac{r_0}{a_+}\bar{t}, \quad \rho = \rho_+\bar{\rho}, \quad p = \gamma p_+\bar{p}, \quad C = a_+\bar{C}, \tag{3.3}$$

where the overhead bar denotes the non-dimensional variable, r_0 is a characteristic length which has been chosen as the initial radius of the piston at $t = 0$.

We take the piston path as a power series in $\bar{t}$

$$\bar{R}_p(\bar{t}) = \bar{r}_0 + R_{p_1}\bar{t} + R_{p_2}\bar{t}^2 + R_{p_3}\bar{t}^3 + \cdots, \tag{3.4}$$

where $R_{p_j} = 0$ for $j > 2$ if the piston acceleration (or deceleration) is constant. We assume that the shock path is also given by a power series

$$\bar{R}(\bar{t}) = \bar{r}_0 + C_1\bar{t} + C_2\bar{t}^2 + C_3\bar{t}^3 + \cdots \tag{3.5}$$

and also that the solution in a small neighbourhood of $\bar{t} = 0$ can be expanded in a Taylor's series of the form

$$\bar{\rho} = \rho_0 + \rho_{11}(\bar{r} - \bar{r}_0) + \rho_{12}\bar{t} + \cdots, \quad \bar{u} = u_0 + u_{11}(\bar{r} - \bar{r}_0) + u_{12}\bar{t} + \cdots,$$
$$\bar{p} = p_0 + p_{11}(\bar{r} - \bar{r}_0) + p_{12}\bar{t} + \cdots, \tag{3.6}$$

where $\bar{r}_0$ is the non-dimensional value of r_0, and ρ_0, u_0, p_0 are the limiting values of the variables $\bar{\rho}, \bar{u}, \bar{p}$ as we approach the shock front from the piston at $\bar{t} = 0$.

Differentiating (3.4) and (3.5) with respect to $\bar{t}$, we get the series expansion for the piston velocity and the shock velocity respectively. Given the coefficients $R_{p_j}, j = 1, 2 \ldots,$ we need to find the coefficients in (3.6) and those for the shock path in (3.5). This is quite straightforward, but involves complex algebraic manipulations. Substituting the expansions (3.6) into the gas-dynamic equations and equating the coefficients of various powers of $\bar{r}$ and $\bar{t}$, we get an undetermined system of linear algebraic equations for the coefficients appearing in (3.6).

To complete this system, we use the series expansion of various quantities appearing in the boundary condition at the piston, namely the equation (3.2) and use the Rankine–Hugoniot condition on the piston path (3.4) to obtain the following set of linear equations for the determination of ρ_{11}, u_{11} and C_2,

$$L_1u_{11} + M_1\rho_{11} + N_1C_2 + P_1 = 0, \tag{3.7}$$
$$L_2u_{11} + M_2\rho_{11} + N_2C_2 + P_2 = 0, \tag{3.8}$$
$$L_3u_{11} + M_3\rho_{11} + N_3C_2 + P_3 = 0, \tag{3.9}$$

where

$$L_1 = 2\rho_0(R_{p_1} - C_1), \ M_1 = (R_{p_1} - C_1)^2, \ N_1 = 2(\rho_0 - 1),$$
$$P_1 = (j/\bar{r}_0)\rho_0R_{p_1}(R_{p_1} - C_1) + 2\rho_0R_{p_2},$$

$$
\begin{aligned}
L_2 &= \gamma p_0 + 3\rho_0(R_{p_1} - C_1)^2, \quad M_2 = (R_{p_1} - C_1)^3, \quad N_2 = 4\rho_0(R_{p_1} - C_1) + 4C_1,\\
P_2 &= 4(R_{p_1} - C_1)\rho_0 R_{p_2} + (j\gamma/\bar{r}_0)p_0 R_{p_1} + (j/\bar{r}_0)\rho_0 R_{p_1}(R_{p_1} - C_1)^2,\\
L_3 &= \gamma^2 p_0/(\gamma - 1) + 3\rho_0(R_{p_1} - C_1)^2/2 - \rho_0(1/(\gamma - 1) + C_1^2/2),\\
M_3 &= (R_{p_1} - C_1)^3/2 - (R_{p_1} - C_1 - 1)(1/(\gamma - 1) + C_1^2/2), \quad N_3 = 2\rho_0 R_{p_1},\\
P_3 &= 2\rho_0 R_{p_2}(2\gamma - 1)(C_1 - R_{p_1})/(\gamma - 1) + (j\gamma/\bar{r}_0)p_0 R_{p_1}\\
&\quad + (j\rho_0 R_{p_1}/\bar{r}_0)\{1/(\gamma - 1) + C_1^2/2 - (C_1 - R_{p_1})^2)/2\}.
\end{aligned} \tag{3.10}
$$

Solving the above system of algebraic equations, we get the values of the required coefficients in (3.6). Hence, finally the initial conditions at $\bar{t} = 0$, for π_0 and π_1 in non-dimensional form is obtained as

$$
\bar{\pi}_0 = R_{p_1}/(R_{p_1} - C_1), \quad \bar{\pi}_1 = (\bar{g}d_{11})^{-1}(-\rho_{11} + jC_1\bar{d}_{10}\bar{\pi}_0/\bar{r}_0), \tag{3.11}
$$

where

$$
C_1 = R_{p_1}(2 - 2\mu^2)^{-1} + (1 + R_{p_1}^2(2 - 2\mu^2)^{-2})^{1/2} \tag{3.12}
$$

which is obtained using Prandtl's relation in the present case of purely radial flow (see [1], p. 425); $\bar{d}_{10}$ and $\bar{d}_{11}$ are the non-dimensional forms of d_{10} and d_{11} respectively, and $\mu^2 = (\gamma - 1)/(\gamma + 1)$. The detailed derivations are available in [14]. We note that the additional term $jC_1\bar{d}_{10}\pi_0/\bar{r}_0$ on the right hand side of (3.12) arising purely due to the geometry of the shock front causes an additional deceleration which accounts for the usual geometric attenuation for the curved shock fronts. By putting $j = 0$ in our equations, it is easily seen that the problem reduces to that of a plane shock (see [6]). In this case it is obvious that the terms P_j, $j = 1, 2, 3$ vanish if $R_{p_2} = 0$, which in turn implies that ρ_{11} and consequently π_1 at $t = 0$ also vanish. Physically, it corresponds to the case of a plane piston moving with a uniform velocity giving rise to a shock of uniform strength.

Thus, it is seen that the initial conditions for the equations (2.17), (2.24) and (2.27) are completely determined in terms of coefficients appearing in the power series expansion of the piston path (3.4). It is also observed that the initial condition for π_0 depends on the piston velocity and the effects of any perturbation (i.e. acceleration or deceleration) in the uniform piston velocity are contained in the initial value for π_1.

4. Approximate solution for weak shock propagation

Some interesting well known results for weak shock propagation can be obtained by assuming that the shock strength is of the order of a small quantity ϵ, i.e. we assume that

$$
\pi_0 = \sum_{j=1}^{\infty} \epsilon^j \pi_0^{(j)}(t), \quad \pi_1 = \sum_{j=0}^{\infty} \epsilon^j \pi_1^{(j)}(t). \tag{4.1}
$$

It is to be noted that the expansion for π_0 starts with the first power of ϵ whereas that for π_1 starts with a constant term. We further assume that $R(t)$ can also be expanded as

$$
R(t) = \sum_{j=0}^{\infty} \epsilon^j R^{(j)}. \tag{4.2}
$$

Substituting the above expansions into the equations of NTSD, and retaining terms only up to first order, we obtain the following set of equations for $\pi_0^{(1)}, \pi_1^{(0)}, R^{(0)}$ and $R^{(1)}$

$$\frac{\mathrm{d}\pi_0^{(1)}}{\mathrm{d}t} = -\pi_0^{(1)}\left(-\frac{\gamma+1}{4\rho_+}\pi_1^{(0)} + \frac{ja_+}{2r}\right), \tag{4.3}$$

$$\frac{\mathrm{d}\pi_1^{(0)}}{\mathrm{d}t} = \frac{\gamma+1}{2\rho_+}(\pi_1^{(0)})^2 - \frac{ja_+}{2r}\pi_1^{(0)}, \tag{4.4}$$

$$\frac{\mathrm{d}R^{(0)}}{\mathrm{d}t} = a_+, \quad \frac{\mathrm{d}R^{(1)}}{\mathrm{d}t} = -a_+\left(\frac{\gamma+1}{4}\right)\pi_0^{(1)}. \tag{4.5}$$

The equations (4.3) and (4.4) can be exactly integrated subject to initial conditions for $\pi_0^{(0)}$ and $\pi_1^{(0)}$ at $t = 0$, say

$$\pi_0^{(1)} = \pi_{00}, \quad \pi_1^{(0)} = \pi_{10}. \tag{4.6}$$

We note that the expansion (4.2) does not hold near the centre as $r \to 0$. Hence we assume that $R^{(0)} \neq 0$ (i.e. the shock front has a finite radius at $t = 0$ which indeed is the case with conventional explosive charges), say $R^{(0)} = r_0$ at the initial instant. Then from (4.5), we have

$$r \approx R^{(0)} = r_0 + a_+t. \tag{4.7}$$

Plane case: In this case, $j = 0$ and the equations (4.5) are not required. The solution to the equations (4.3) and (4.4) when integrated with the initial conditions (4.6) are

$$\pi_1^{(0)} = \pi_{10}(1 - (\gamma+1)\pi_{10}t/(2\rho_+))^{-1}, \tag{4.8}$$

$$\pi_0^{(1)} = \pi_{00}(1 - (\gamma+1)\pi_{10}t/(2\rho_+))^{-1/2}. \tag{4.9}$$

Cylindrical case: In this case (i.e. when $j = 1$) the solutions to the equations (4.3) and (4.4) assume the following form

$$\pi_1^{(0)}(t) = \frac{a_+\rho_+\pi_{10}r_0^{1/2}}{(\rho_+ + (\gamma+1)r_0\pi_{10})(r_0+a_+t)^{1/2} - (\gamma+1)(r_0+a_+t)\pi_{10}r_0^{1/2}}, \tag{4.10}$$

$$\pi_0^{(1)}(t) = \pi_{00}\{(r_0+a_+t)((\rho_+ + (\gamma+1)r_0\pi_{10}) - (\gamma+1)\pi_{10}r_0^{1/2}(r_0+a_+t)^{1/2}\}^{-1/2}. \tag{4.11}$$

Spherical case: In this case (i.e. $j = 2$), the solutions for (4.3) and (4.4) are given by

$$\pi_1^{(0)}(t) = \frac{2a_+\rho_+\pi_{10}r_0}{(r_0+a_+t)(2a_+\rho_+ + (\gamma+1)\pi_{10}r_0\log r_0 - (\gamma+1)\pi_{10}r_0\log(r_0+a_+t))}, \tag{4.12}$$

$$\pi_0^{(1)}(t) = \frac{\pi_{00}}{(r_0+a_+t)}(2a_+\rho_+ + (\gamma+1)r_0\pi_{10}\log r_0 - (\gamma+1)r_0\pi_{10}\log(r_0+a_+t))^{-1/2}. \tag{4.13}$$

Critical time: It is seen that if $\pi_{10} > 0$, then the solutions given by (4.8)–(4.13) cannot be continued beyond a time t_c (called the critical time). It is also seen that as $t \to t_c$, $\pi_1^{(0)}$ and $\pi_0^{(1)}$ approach infinitely large values in each of the above cases. In fact, the weak shock assumption breaks down before these quantities approach infinity and t_c is an indication of this. The critical time t_c in the above three cases are given by

$$\text{Plane case:} \quad t_c = 2\rho_+/((\gamma+1)\pi_{10}), \tag{4.14}$$

$$\text{Cylindrical case:} \quad t_c = a_+^{-1}\{(a_+\rho_+/\pi_{10} + (\gamma+1)r_0)^2/((\gamma+1)^2 r_0) - r_0\}, \tag{4.15}$$

$$\text{Spherical case:} \quad t_c = a_+^{-1}\{\exp(2a_+\rho_+/((\gamma+1)\pi_{10}r_0) + \log r_0) - r_0\}. \tag{4.16}$$

In all the above three cases, we observe that

(i) there is a positive value for t_c for all positive values of π_{10}, which corresponds to the case of accelerating piston,

(ii) there is no finite value for t_c when π_{10} is negative, (i.e. the solutions (4.8)–(4.13) can be continued for all times for all negative values of π_{10}).

We also note that the case $\pi_{10} < 0$ corresponds to the case when the slope of the density versus spatial coordinate curve is positive, which implies that no positive finite value for t_c exists and hence the solution can be continued for all times (see [10]).

Comparison with exact results for decay of weak shocks: Taking the limit as $t \to \infty$ in (4.9), (4.11) and (4.13), so that the terms independent of t can be ignored, it is seen that the shock strength decays as $t^{-1/2}$, $t^{-3/4}$ and $t^{-1}(\log t)^{-1/2}$ for the cases of plane, cylindrical and spherical shocks respectively. Thus, the decay rule for weak shocks from NTSD agrees with classical results for the asymptotic decay for the cylindrical and spherical waves (see Landau [5], Whitham [19] and also Grinfeld [2]).

5. Results and discussions

The system of ordinary differential equations (2.17) and (2.24) (with $\pi_2 = 0$) and (2.27) are solved using Runge–Kutta–Gill method. In figure 1, the case of accelerating

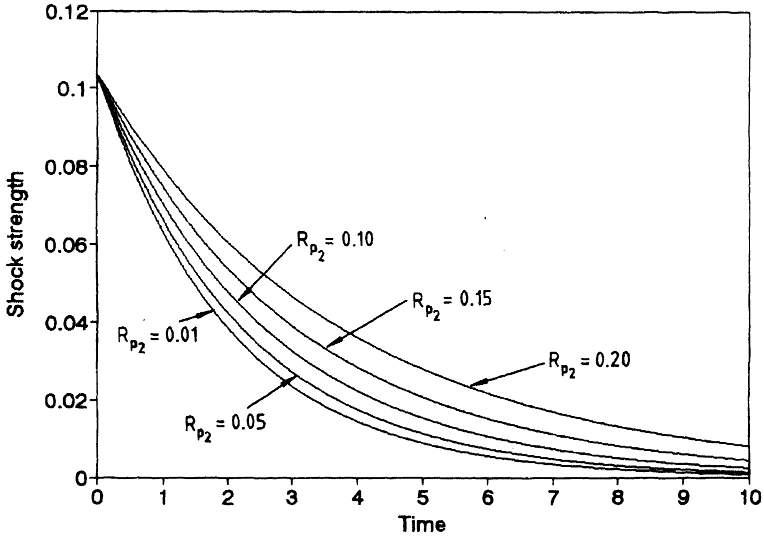


Figure 1. Decay of a cylindrical shock, originating from an initial piston velocity $R_{p1} = 0.10$ and with indicated values of accelerations R_{p2}.

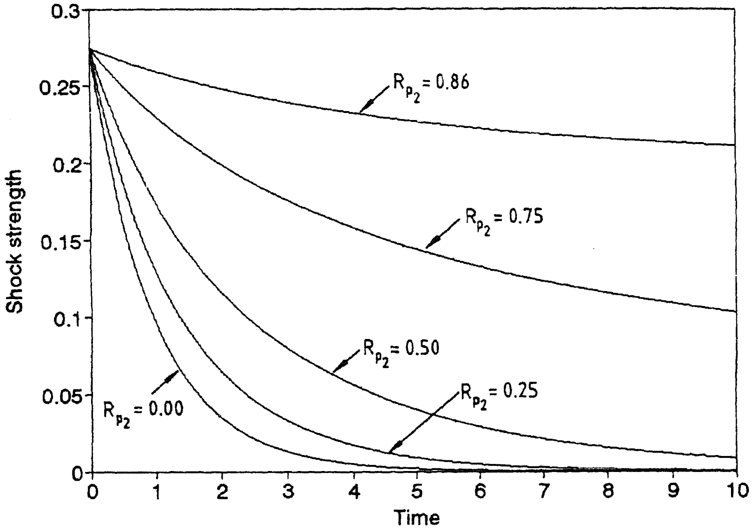


Figure 2. Decay of a spherical shock originating from an initial piston velocity $R_{p1} = 0.25$ and with indicated values of accelerations.

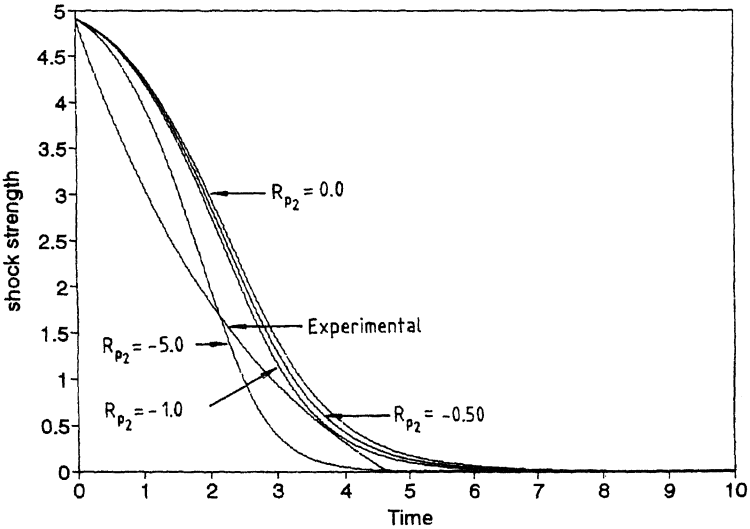


Figure 3. Decay of a strong spherical shock, originating from an initial velocity $R_{p1} = 15.0$.

cylindrical piston is shown; the initial piston velocity R_{p_1} is chosen as 0.10 and the value chosen for the acceleration R_{p_2} are 0.01, 0.05, 0.10, 0.15 and 0.20 respectively. In figure 2 the case of accelerating spherical piston is shown corresponding to $R_{p_1} = 0.25$ and $R_{p_2} =$ 0.00, 0.25, 0.50, 0.75 and 0.86 respectively. The attenuation of the shock solely due to th geometrical effects is obvious in these curves. It is also observed (as in figure 2) that t maintain the initial strength of the shock (i.e. to overcome the effects of geometrica attenuations), a considerable amount of constant acceleration is required.

In figure 3, attenuation of a very strong spherical shock corresponding to $R_{p1} = 15.$ with varying amounts of decelerations $R_{p_2} = 0.00, -0.50, -1.00$ and -5.00 is show

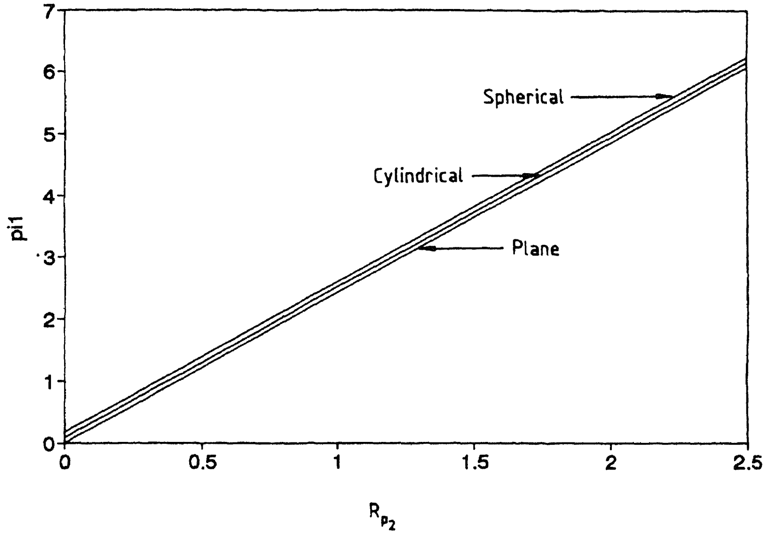


Figure 4. Initial values of π_1 at $t = 0$ as functions of the piston accelerations R_{p2} for the plane, cylindrical and spherical cases.

Since, in an actual blast phenomenon, the strong spherical or cylindrical shock front is attenuated due to geometrical effects as well as the rarefaction waves following it, the decelerating spherical or cylindrical piston problem may serve as a simplified and approximate model to represent the actual phenomenon. For comparison, we have also plotted the time decay curve for 1000 kg TNT at a distance of 1 m from the point of detonation (see Kinney and Graham [4]). We have chosen this particular example as the particle velocity in this case is comparable with the piston velocity under consideration at $t = 0$. For this problem $r_0 = 0.548\,\mathrm{m}$ and $\bar{t} \approx 620t$. Since in an actual blast, the decay pattern depends on a number of factors such as the chemical composition of the explosive, its packing density etc. which are not included in our mathematical formulation, it may explain the deviation of NTSD results from the experimental curves.

As described in § 3, the value of $\pi_1|_{t=0}$ depends on the piston acceleration (or deceleration) R_{p_2}. For a typical case of $R_{p_1} = 0.25$, we have plotted $\pi_1|_{t=0}$ against R_{p_2} in figure 4 for the plane, cylindrical and spherical cases. The relationship is nearly linear. It is also seen that $\pi_1|_{t=0} = 0$ for the plane case whereas it has small nonzero positive values for the cases of curved pistons.

As indicated in §§ 3 and 4, the value of the critical time t_c would depend upon the piston acceleration. We have tabulated the values of t_c for a typical case of $R_{p_1} = 0.25$ and R_{p_2} varying from 0.25 to 5.0 in table 1. It is observed that with gradual increase in R_{p_2}, the plane shock reaches the strong shock limit (i.e. $\pi_0 \to 5.0$) and $\pi_1 \to \infty$ as $t \to t_c$ in almost all the cases. The curved shocks behave in a different way: they continue to decay in the presence of comparatively smaller values of piston acceleration (as they do in its absence). It is seen that there is a threshold value for R_{p_2} below which the curved shocks decay. There is another threshold value for R_{p_2} up to which the strength of a curved shock reaches a constant value in a finite time. The entry 'con' in the table 1 refers to this value and t is the time at which this constant shock strength is reached. Beyond this second threshold value for R_{p_2}, the shock strength grows until it attains the strong shock limit and $\pi_1 \to \infty$ at $t = t_c$ where the NTSD algorithm breaks down.

Table 1. Critical times t_c for $R_{p1} = 0.25$.

R_{p2}	Plane	Cylindrical	Spherical
0.25	con^{1}	sd	sd
0.50	26.3960	con^{2}	sd
0.75	17.5791	con^{3}	con^{8}
1.00	13.1743	con^{4}	con^{9}
1.25	10.5197	con^{5}	con^{10}
1.50	8.7600	con^{6}	con^{11}
1.75	7.5052	con^{7}	con^{12}
2.00	6.5601	175.4784	con^{13}
2.25	5.8251	25.4558	con^{14}
2.50	5.2401	15.7289	con^{15}
2.75	4.7601	11.6196	con^{16}
3.00	4.3551	9.2849	con^{17}
3.25	4.0201	7.7552	con^{18}
3.50	3.7301	6.6701	con^{19}
3.75	3.4751	5.8601	44.0306
4.00	3.2551	5.2301	17.3091
4.25	3.0601	4.7201	11.7045
4.50	2.8900	4.3051	9.0350
4.75	2.7350	3.9551	7.4201
5.00	2.5950	3.6601	6.3301

where sd: shock decays, con^{1}: $\pi_0 = 5.0000$ at $t = 52.8258$, con^{2}: $\pi_0 = 0.39028$ at $t = 54.1351$, con^{3}: $\pi_0 = 1.08578$ at $t = 37.6291$, con^{4}: $\pi_0 = 1.78331$ at $t = 35.5953$, con^{5}: $\pi_0 = 2.48500$ at $t = 41.0972$, con^{6}: $\pi_0 = 3.19524$ at $t = 49.4776$, con^{7}: $\pi_0 = 3.92700$ at $t = 71.6121$, con:^{8}: $\pi_0 = 0.04452$ at $t = 185.5282$, con^{9}: $\pi_0 = 0.39282$ at $t = 29.5517$, con^{10}: $\pi_0 = 0.74192$ at $t = 24.1956$, con^{11}: $\pi_0 = 1.09205$ at $t = 19.3595$, con^{12}: $\pi_0 = 1.44375$ at $t = 19.8696$, con^{13}: $\pi_0 = 1.79763$ at $t = 21.0199$, con^{14}: $\pi_0 = 2.15451$ at $t = 20.2747$, con^{15}: $\pi_0 = 2.51575$ at $t = 20.7798$, con^{16}: $\pi_0 = 2.88369$ at $t = 24.0205$, con^{17}: $\pi_0 = 3.26236$ at $t = 30.9670$, con^{18}: $\pi_0 = 3.65961$ at $t = 35.0955$, con^{19}: $\pi_0 = 4.10130$ at $t = 43.9806$.

Acknowledgements

The authors are thankful to Prof. Phoolan Prasad, Department of Mathematics, IISc, Bangalore for valuable discussions and Dr N K Srinivasan, Director, CASSA, Bangalore for giving permission to publish this paper.

References

[1] Courant R and Friedrichs K O, *Supersonic flows and shock waves* (Interscience Publishers) (1948)

[2] Grinfeld M A, Ray method for calculating the wave front intensity in non-linear elastic material, *PMM J. App. Math. Mech.* **42** (1978) 958–977

[3] Holt M, *Numerical Methods in Fluid Dynamics* (Berlin, Heidelberg: Springer-Verlag) (1984)

[4] Kinney G F and Graham K J, *Explosive Shock Waves in Air*, 2nd ed. (Berlin, Heidelberg: Springer-Verlag) (1985)

[5] Landau L D, On shock waves at large distances from the place of their origin, *Sov. J. Phys.* **9** (1945) 496–500

[6] Lazarev M P, Prasad P and Singh S K, An approximate solution of one dimensional piston problem, *ZAMP* **46** (1995) 752–771

[7] Maslov V P, Propagation of shock waves in an isotropic non-viscous gas, *J. Sov. Math.* **13** (1980) 119–163

[8] Peyret R and Taylor T D, *Computational Methods in Fluid Flow* (Berlin, Heidelberg: Springer-Verlag) (1983)

[9] Prasad P, Kinematics of a multi-dimensional shock of arbitrary strength in an ideal gas, *Acta Mechanica* **45** (1982) 163–176

[10] Prasad P, *Propagation of a curved shock and nonlinear ray theory*, Pitman Research Notes in Mathematics Series 292 (Harlow: Longman Scientific and Technical) (1993)

[11] Ravindran R and Prasad P, A new theory of shock dynamics. Part I: Analytic considerations, *Appl. Math. Lett.* **3** (1990) 77–81; Part II: *Numerical Results* **3** (1990) 107–109

[12] Sedov L I, *Similarity and Dimensional Methods in Mechanics* (Moscow: Mir Publishers) (1982)

[13] Singh S K and Singh V P, A note on the new theory of shock dynamics, *Def. Sci. J.* **42** (1992) 103–105

[14] Singh S K, *Some Problems on the Propagation of Plane and Curved Shocks* Ph. D. thesis (Bangalore: Indian Institute of Science) (1997)

[15] Stanyukovich K P, *Unsteady motion of a continuous media* (London: Pergamon Press) (1960)

[16] Taylor G I, The formation of a blast wave by a very intense explosion. I: Theoretical discussion, *Proc. R. Soc.* **A201** (1950) 159–174

[17] Thomas T Y, *Plastic flow and fracture of solids* (New York: Academic Press) (1961)

[18] Whitham G B, A new approach to problems of shock dynamics, Part I: Two-dimensional problems, *J. Fluid Mech.* **2** (1957) 140–171; Part II: Three-dimensional Problems, **5** (1959) 369–386

[19] Whitham G B, *Linear and Nonlinear Waves* (New York: John-Wiley and Sons) (1974)

Proc. Indian Acad. Sci. (Math. Sci.), Vol. 110, No. 1, February 2000, pp. 117–120.

Slow rotation of a sphere with source at its centre in a viscous fluid

SUNIL DATTA and DEEPAK KUMAR SRIVASTAVA

Department of Mathematics and Astronomy, Lucknow University, Lucknow 226 007, India

MS received 10 March 1999; revised 12 October 1999

Abstract. In this note, the problem of a sphere carrying a fluid source at its centre and rotating with slow uniform angular velocity about a diameter is studied. The analysis reveals that only the azimuthal component of velocity exists and is seen that the effect of the source is to decrease it. Also, the couple on the sphere is found to decrease on account of the source.

Keywords. Slow rotation; viscous fluid.

1. Introduction

The problem of determining the couple experienced by axially symmetric bodies, rotating steadily in a viscous and incompressible fluid has engaged the attention of many workers like Jeffery [2], Kanwal [3], Smith [6], Watson [7], and Ram Kissoon [5]. The purpose of this paper is to study slow rotation of a sphere, assumed to be pervious, with a source at its centre. If the strength Q of the source were of the same order as the angular velocity Ω of rotating sphere, the inertia terms could still be neglected and the total flow consists of only the source solution superimposed on the Stokes solution; thus in this case the Stokes drag and couple are not affected by the source. On the other hand if Q is large enough so that the $Q\Omega$ is not negligible, the inertia terms, being non-linear, cannot be altogether omitted; the equation, however, can still be linearized by assuming that the velocity perturbation in the source flow on account of the Stokes flow is small so that the terms containing square of angular velocity can be neglected. This assumption is justifiable at least in the vicinity of the sphere where the Stokes approximation is valid too. The problem corresponds to the problem of Stokes flow past a sphere with source at its centre investigated by Datta [1], the results of which have found application in investigating the diffusiophoresis target efficiency for an evaporating or condensing drop [4].

2. Formulation of the problem

Let us consider a pervious sphere of radius a with source of strength Q at its center generating radial flow field around it in an infinite expanse of incompressible fluid of density ρ and kinematic viscosity ν. The sphere is also made to rotate with small steady angular velocity Ω so that terms of an $O(\Omega^2)$ may be neglected but terms of $O(Q\Omega)$ is retained.

The motion is governed by Navier–Stokes equations and the continuity equation together with no-slip boundary condition

$$\boldsymbol{u} = a\Omega\hat{e}_x x\hat{e}_r, \tag{2.1}$$

on the surface $r = a$, and the condition of vanishing of velocity at far off points,

$$\boldsymbol{u} = 0, \quad \text{at infinity as } r \to \infty. \tag{2.2}$$

It will be convenient to work in spherical polar coordinates (r, θ, ϕ) with x-axis as the polar axis. We non-dimensionalize the space variables by a, velocity $\boldsymbol{u}(v_r, v_\theta, v_\phi)$ by $a\Omega$ and pressure by $\rho\nu\Omega$. Moreover, the symmetry of the problem and the boundary conditions ensure that velocity components $v_r = v_\theta = 0$, and then we may express the velocity vector $\boldsymbol{u}$ as

$$\boldsymbol{u} = \frac{Q}{a^2 r^2} r + a\Omega v_\phi(r, \theta)\hat{e}_\phi \tag{2.3}$$

and pressure as

$$p = \rho\nu\Omega[p_0(r) + p_1(r, \theta)]. \tag{2.4}$$

By using the value (2.3) and (2.4) in Navier–Stokes equation, the azimuthal componant v_ϕ is seen to satisfy the equation

$$\nabla^2 v_\phi - \frac{v_\phi}{r^2 \sin^2\theta} = \frac{s}{r^3}\frac{\partial}{\partial r}(r v_\phi), \tag{2.5}$$

where $s = Q\Omega/\nu a$ is the source parameter.

The above equation is to be solved under the boundary conditions

$$\left(\begin{array}{ll} v_\phi = a\Omega \sin\theta & \text{at } r = 1 \\ \text{and} & \\ v_\phi \to 0 & \text{as } r \to \infty. \end{array}\right. \tag{2.6}$$

3. Solution

We substitute

$$v_\phi = r\omega(r) \sin\theta \tag{3.1}$$

in equation (2.5) and solve the resulting differential equation in $\omega(r)$, using boundary conditions corresponding to (2.6) to get

$$v_\phi = r\sin\theta\left[\frac{s^2}{r^2} - 2\frac{s}{r} + 2 - 2\mathrm{e}^{-g/r}\right][s^2 - 2s + -2\mathrm{e}^{-g}]^{-1}. \tag{3.2}$$

Following limiting values of v_ϕ emerge from (3.2)

Case I. When parameter $s = Q\Omega/\nu a$ is small, we have

$$v_\phi \approx \frac{\sin\theta}{r^2}\left[1 + \frac{s}{4}\left(1 - \frac{1}{r}\right)\right]. \tag{3.3}$$

Case II. When parameter $s = Q\Omega/\nu a$ is large and r is finite, we get

$$v_\phi \approx \frac{\sin\theta}{r^2}\left[1 - \frac{2}{s}(r - 1)\right]. \tag{3.4}$$

Since v_ϕ can not be negative the above result is to be true for $1 \le r < 1 + s/2$.

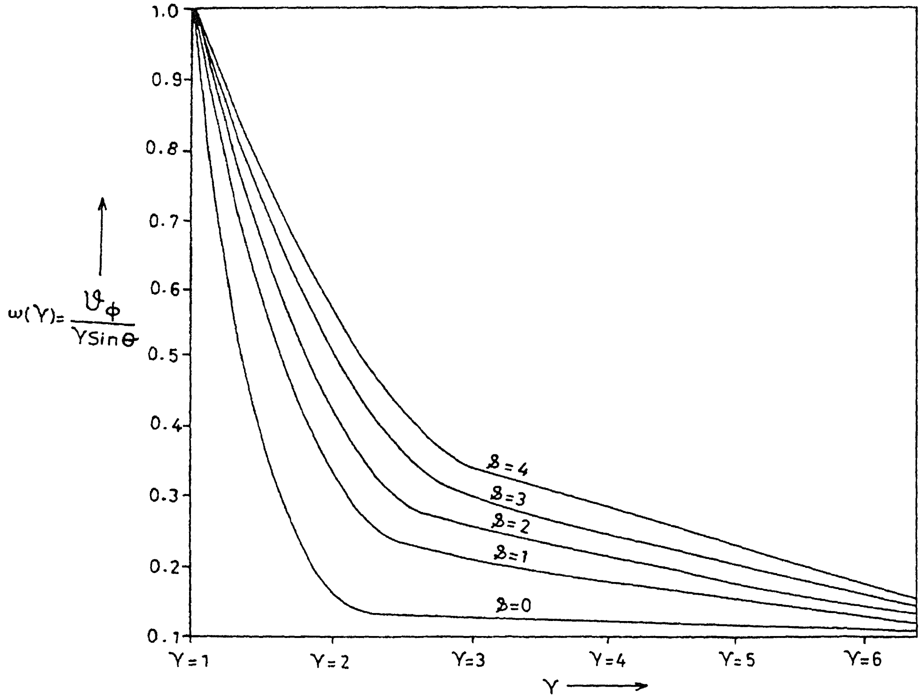


Figure 1. Variation of angular velocity $\omega(r)$ with respect to r and for varions values of parameter s.

4. Couple on the sphere

The couple on the sphere required to maintain the motion is obtained by integration of viscous stress

$$\sigma_{r\phi} = \mu\Omega\left(\frac{\partial v_\phi}{\partial r} - \frac{v_\phi}{r}\right)_{r=1}.$$

Thus, the moment of the required couple is given by

$$\begin{aligned} M &= -\int_{\theta=0}^{\pi} \sigma_{r\phi} R^3 2\pi \sin\theta \mathrm{d}\theta \\ &= \frac{16}{3}\pi a^3 \mu\Omega[s(1-\mathrm{e}^{-s}) - s^2][2 - 2s + s^2 - 2\mathrm{e}^{-s}]^{-1}. \end{aligned} \tag{4.1}$$

Further, when s is small, we get from (4.1)

$$M = 8\pi\mu a^3\Omega\left(1 - \frac{s}{12}\right), \tag{4.2}$$

which provides the classical value $M_0 = 8\pi\mu a^3\Omega$ for $s = 0$. Also for large values of s, we have the approximation as

$$M = \frac{16}{3}\pi\mu a^3\Omega\left(1 + \frac{1}{s}\right). \tag{4.3}$$

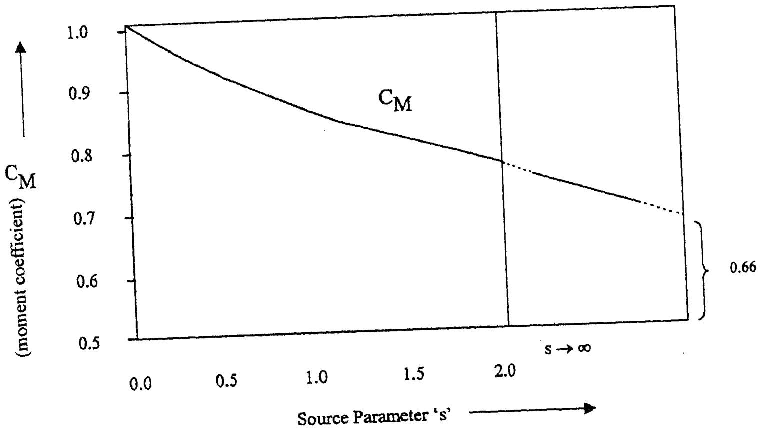


Figure 2. Variation of moment coefficient C_M with respect to source parameter s.

We find that M tends to 2/3 M_0, as $s \to \infty$. The expression for angular velocity $\omega(r) = v_\phi/r \sin\theta$ can be easily obtained from equation (3.2) as

$$\omega(r) = \left[\frac{s^2}{r^2} - 2\frac{s}{r} + 2 - 2e^{-s/r}\right][s^2 - 2s + 2 - 2e^{-s}]^{-1}. \tag{4.4}$$

The general behaviour of angular velocity $\omega(r)$ with respect to r and for various values of parameter s has been shown in figure 1. It may be concluded that flow gets dampened as the source strength increases.

The variation of moment coefficient $C_M = M/M_0$ with source parameter $s = Q\Omega/\nu a$, where M_0 is the moment for $s = 0$, has been shown in figure 2.

Figure (2) clearly shows that the effect of source is to reduce the moment ultimately to two third of its value in the absence of the source.

References

[1] Datta S, Stokes flow past a sphere with a source at its centre. *Math. Vesnik* **10(25)** (1973) 227–229

[2] Jeffry G B, Steady rotation of a solid of revolution in a viscous fluid. *Proc. London Math. Soc.* **14** (1955) 327–38

[3] Kanwal R P, Slow steady rotation of axially symmetric bodies in a viscous fluid. *J. Fluid Mech.* **10** (1960) 17–24

[4] Placek T D and Peters L K, A hydrodynamic approach to particle target efficiency in the presence of diffusiophoresis. *Aerospace J.* **11** (1980) 521–533

[5] Ram Kissoon H, A Slip flow problem. *J. Math. Sci (Calcutta)* **8(1)** (1997) 23–27

[6] Smith S H, The rotation of two circular cylinders in a viscous fluid. *Mathematika*, **38(1)** (1991) 63–66

[7] Watson E J, Slow viscous fluid past two rotating cylinders, *Q. J. Mech. Appl. Math.* **49(2)** (1996) 195–216

Proc. Indian Acad. Sci. (Math. Sci.), Vol. 110, No. 2, May 2000, pp. 121–132.
© Printed in India

How to recover an L-series from its values at almost all positive integers. Some remarks on a formula of Ramanujan

CHRISTOPHER DENINGER

WWU, Mathematisches Institut, Einsteinstrasse 62, D-48149 Munster, Germany

MS received 25 August 1999; revised 15 January 2000

Abstract. We define a class of analytic functions which can be obtained from their values at almost all positive integers by a canonical interpolation procedure. All the usual L-functions belong to this class which is interesting in view of the extensive investigations of special values of motivic L-series. A number of classical contour integral formulas appear as particular cases of the interpolation scheme. The paper is based on a formula of Ramanujan and results of Hardy. An approach to the problem via distributions is also presented.

Keywords. Interpolation formulas; analytic functions; contour integrals; special values; L-functions

1. Introduction

The purpose of this note is to answer a question, Mazur asked me: Is there an interpolation scheme allowing to recover a complex L-series from its values at *almost all* positive integers? This is interesting for example in view of the extensive investigations of special values of motivic L-series in the last decades, culminating in the Bloch–Kato conjectures [BK]. Note that p-adic L-functions are determined by their values at these points because the set in question is dense in $\mathbb{Z}_p$.

It turns out that modifying one of Ramanujan's favourite formulas one gets a satisfactory interpolation procedure for a class of analytic functions which in particular comprises all Dirichlet series and hence all L-series. Incidentally the classical representations of certain zeta- and L-functions as contour integrals are particular instances of the interpolation scheme.

Ramanujan did not specify exactly to which functions his formula applied. A useful class $\mathcal{F}_H$ was singled out however by Hardy in his commentary on Ramanujan's work [H], ch. XI. We introduce a universal class $\mathcal{F}$ of interpolizable functions which is essentially canonical. Hardy's result then implies that $\mathcal{F}_H \subset \mathcal{F}$. Apart from the general setup and a number of examples we also give a short distribution theoretic proof of a special case of Hardy's result. This uses Schwartz' extension of the Paley–Wiener theorem to distributions with compact support.

It should be emphasized that this note is essentially a commentary on one aspect of the work of Hardy and Ramanujan.

2. Preliminaries

To a sequence of complex numbers $\mathbf{a} = (a_\nu)_{\nu \geq \nu_0}$, $\nu_0 \in \mathbb{Z}$ we associate the Laurent series:

$$\psi = \psi_{\mathbf{a}}(x) := \sum_{\nu \geq \nu_0} a_\nu x^\nu .$$

Extend **a** to a sequence indexed by the integers by setting $a_\nu = 0$ for $\nu < \nu_0$. We call ψ 'good' if the following conditions are satisfied:

$$\psi \text{ converges in a punctured neighborhood of the origin,} \tag{1}$$

$$\psi \text{ has a holomorphic continuation to } U' \text{ where } U \text{ is a neighborhood of } (-\infty, 0] \text{ and } U' = U \setminus \{0\}. \tag{2}$$

For some $\alpha \in \mathbb{R}$ we have

$$|\psi(z)| = O(|z|^\alpha) \text{ as } |z| \to \infty \text{ in } U'. \tag{3}$$

If ψ is good, the contour integral

$$I(\psi)(s) := \frac{1}{2\pi i} \int_{-\infty}^{(0+)} z^{-s} \psi(z) \frac{\mathrm{d}z}{z} \tag{4}$$

defines a holomorphic function of s in $\mathrm{Re}\, s > \alpha$ for any α as in condition (3). Here the integration is along any path within $U \setminus (-\infty, 0]$ starting at $-\infty$, encircling the origin counterclockwise and returning to $-\infty$.

The power z^{-s} is defined via $\log z = \log |z| + i \operatorname{Arg} z$ where $-\pi < \operatorname{Arg} z \leq \pi$. Note that $I(\psi)$ does not depend on the choice of the path.

Lemma 1.1. *In the situation of* (4) *we have:*

$$I(\psi)(\nu) = a_\nu \quad \text{for } \nu > \alpha.$$

Proof. For $s = \nu > \alpha$ since $z^{-\nu}$ is single valued the integral reduces to

$$I(\psi)(\nu) = \frac{1}{2\pi i} \oint_{|z|=\varepsilon} z^{-\nu} \psi(z) \frac{\mathrm{d}z}{z} = \mathrm{Res}_{z=0} \left(z^{-\nu} \psi(z) \frac{\mathrm{d}z}{z} \right) = a_\nu$$

by taking ε sufficiently small. □

We need another consequence of the residue theorem:

Lemma 1.2. *A rational function ψ of degree d with no poles on* $(-\infty, 0)$ *is good with $\alpha = d$ and we have:*

$$I(\psi)(s) = - \sum_{a \in \mathbb{C} \setminus (-\infty, 0]} \mathrm{Res}_a \left(z^{-s} \psi(z) \frac{\mathrm{d}z}{z} \right) \quad \text{in } \mathrm{Re}\, s > d.$$

Finally we require for later use.

Lemma 1.3. *Assume that $\psi = \sum_{\nu \geq \nu_0} a_\nu x^\nu$ is good with some $\alpha < \nu_0$ in condition* (3). *Then we have*

$$I(\psi)(s) = - \frac{\sin \pi s}{\pi} M(\psi(-x))(-s) \quad \text{for } \alpha < \mathrm{Re}\, s < \nu_0 .$$

Here

$$MF(s) = \int_0^\infty x^s F(x) \frac{\mathrm{d}x}{x}$$

is the Mellin transform on $\mathbb{R}_+^*$.

Proof. For $\operatorname{Re} s > \alpha$ and every $\varepsilon > 0$ small enough we have:

$$\begin{aligned} I(\psi)(s) &= \frac{1}{2\pi i} \int_{-\infty}^{-\varepsilon} \mathrm{e}^{-s(\log|x| - i\pi)} \psi(x) \frac{\mathrm{d}x}{x} + \frac{1}{2\pi i} \int_{-\varepsilon}^{-\infty} \mathrm{e}^{-s(\log|x| + i\pi)} \psi(x) \frac{\mathrm{d}x}{x} \\ &\quad + \frac{1}{2\pi i} \oint_{|z|=\varepsilon} z^{-s} \psi(z) \frac{\mathrm{d}z}{z}, \\ &= -\frac{\sin \pi s}{\pi} \int_\varepsilon^\infty x^{-s} \psi(-x) \frac{\mathrm{d}x}{x} + \frac{1}{2\pi i} \oint_{|z|=\varepsilon} z^{-s} \psi(z) \frac{\mathrm{d}z}{z}. \end{aligned}$$

Since $|\psi(z)| = O(|z|^{\nu_0})$ as $|z| \to 0$ we have the estimate

$$\left| \oint_{|z|=\varepsilon} \right| \leq c\varepsilon^{\nu_0 - \operatorname{Re} s} \quad \text{and hence} \quad \lim_{\varepsilon \to 0} \oint_{|z|=\varepsilon} = 0$$

for $\operatorname{Re} s < \nu_0$. Hence the formula. □

Remark. The theory of the Mellin transform is well developed. In [I], Theorem 3.1 for example two function spaces are defined which are in bijection via the Mellin transform. Together with Lemma 1.3, Igusa's result leads to information about the interpolation functional $I(\psi)$. Unfortunately the class of functions ψ to which we want to apply I in the next sections is quite different from the one that can be treated in this way.

3. Interpolation

We can now set up the interpolation scheme. Consider the $\mathbb{C}$-algebra:

$$\mathcal{A}' := \{\text{sequences } (a_\nu) \text{ defined from some } \nu_0 \text{ onwards}\} / \sim$$

where $\mathbf{a} = (a_\nu)_{\nu \geq \nu_0} \sim \mathbf{a}' = (a'_\nu)_{\nu \geq \nu'_0}$ iff $a_\nu = a'_\nu$ for all $\nu \gg 0$.
If sequences $\mathbf{a}, \mathbf{a}'$ are equivalent, then $\psi_{\mathbf{a}}$ is good iff $\psi_{\mathbf{a}'}$ is good. Hence we can define:

$$\mathcal{A} := \{[\mathbf{a}] \in \mathcal{A}' \mid \psi_{\mathbf{a}} \text{ is good}\}.$$

On the other hand let $\mathcal{F}'$ be the $\mathbb{C}$-vector space of holomorphic functions ϕ defined on some half plane $\operatorname{Re} s > \beta$ s.t. $\psi(x) = \sum_{\nu > \beta} \phi(\nu) x^\nu$ is good. Here the summation is over all integers $\nu > \beta$. By the principle of analytic continuation we may identify functions in $\mathcal{F}'$ if they agree for $\operatorname{Re} s \gg 0$.

Theorem 2.1. *I defines a linear 'interpolation' map*

$$I : \mathcal{A} \longrightarrow \mathcal{F}' \text{ via } I([\mathbf{a}]) = I(\psi_{\mathbf{a}}).$$

It has the 'special-values map'

$$S : \mathcal{F}' \longrightarrow \mathcal{A}, S(\phi) = [(\phi(\nu))_{\nu \gg 0}]$$

as a left-inverse:

$$S \circ I = \mathrm{id}.$$

Proof. For any sequence $\mathbf{a} = (a_\nu)_{\nu \geq \nu_0}$ such that $\psi_{\mathbf{a}}$ is good the function $\phi(s) = I(\psi_{\mathbf{a}})(s)$ is holomorphic in some half plane $\operatorname{Re} s > \beta$. By Lemma 1.1 it has the property that $(\phi(\nu))_{\nu > \beta} \sim \mathbf{a}$. Hence $\psi(x) = \sum_{\nu > \beta} \phi(\nu) x^\nu$ is good as well and thus $\phi \in \mathcal{F}'$.

If $\mathbf{a} \sim \mathbf{a}'$ then $\psi_{\mathbf{a}} - \psi_{\mathbf{a}'} \in \mathbb{C}[x, x^{-1}]$. By Lemma 1.2 we therefore have $I(\psi_{\mathbf{a}}) = I(\psi_{\mathbf{a}'})$ in $\mathcal{F}'$. Thus the interpolation map is well defined. As we have seen

$$(I(\psi_{\mathbf{a}})(\nu))_{\nu > \beta} \sim \mathbf{a}$$

and hence $S \circ I = \mathrm{id}$ on $\mathcal{A}$. □

We now define:

$$\mathcal{F} := \operatorname{Im} I \subset \mathcal{F}'.$$

Then I and S define mutually inverse $\mathbb{C}$-linear isomorphisms

$$\mathcal{A} \underset{S}{\overset{I}{\rightleftarrows}} \mathcal{F}. \tag{5}$$

This is clear since I was injective having a left-inverse and we have made it surjective.

By construction the functions in $\mathcal{F}$ have the property that they are uniquely determined by their values on any set of integers of the form $\{\nu | \nu \geq \nu_0\}$. Moreover given these values for $\nu \geq \nu_0$ there is an explicit formula for the function, valid in some half plane $\operatorname{Re} s > \beta$.

Note that we have a canonical projector:

$$P = I \circ S : \mathcal{F}' \longrightarrow \mathcal{F}, \; P^2 = P. \tag{6}$$

In these terms we have:

PROPOSITION 2.2

(1) *For $A, A' \in \mathcal{A}$ form $A \cdot A' \in \mathcal{A}'$. If $A \cdot A' \in \mathcal{A}$, then $I(A \cdot A') = P(I(A) \cdot I(A'))$ in $\mathcal{F}$.*
(2) *S and I are equivariant with respect to the $\mathbb{Z}$-action by shift.*

Proof. (1) If $A \cdot A' \in \mathcal{A}$ then $I(A) \cdot I(A') \in \mathcal{F}'$ since $(I(A) \cdot I(A'))(\nu) = I(A)(\nu) \cdot I(A')(\nu) = a_\nu \cdot a'_\nu$ for $\nu \gg 0$ where $(a_\nu)_{\nu \geq \nu_0}, (a'_\nu)_{\nu \geq \nu'_0}$ are representatives of A, A'. Hence $S(I(A) \cdot I(A')) = A \cdot A'$. Applying I gives the assertion. (2) Shift by one acts on $\mathcal{A}'$ by $T[(a_\nu)] = [(a_{\nu+1})]$. The corresponding ψ is $x^{-1}\psi_{\mathbf{a}}$ which is again good. Hence the shift acts on $\mathcal{A}$ and by a similar argument also on $\mathcal{F}'$. The rest is clear. □

Remark. There is a convolution product for sequences but it does not pass to $\mathcal{A}$.

Before we incorporate the Hardy–Ramanujan theory into the picture let us give some examples. For a sequence $\mathbf{a}$ with $[\mathbf{a}]$ in $\mathcal{A}$ we set $I(\mathbf{a}) := I([\mathbf{a}])$.

Example 2.3. For $\lambda \in \mathbb{C}^*$ consider $\mathbf{a} = (\lambda^{-\nu})_{\nu \geq 0}$. Then if $\lambda \notin (-\infty, 0)$ the class of $\mathbf{a}$ is in $\mathcal{A}$ and $I(\mathbf{a}) = \lambda^{-s}$ where $\arg \lambda \in (-\pi, \pi]$. In particular $\phi(s) = \lambda^{-s} \in \mathcal{F}$. The functions λ^{-s} defined using different normalizations of $\arg \lambda$ lie in $\mathcal{F}'$ and are mapped via P to the principal one.

Proof. For $\lambda \notin (-\infty, 0)$ the function

$$\psi(x) = \psi_{\mathbf{a}}(x) = \sum_{\nu=0}^{\infty} \lambda^{-\nu} x^{\nu} = \frac{1}{1 - \lambda^{-1} x}; \ |x| < \lambda$$

is good in our sense. By Lemma 1.2 we have

$$I(\psi)(s) = -\operatorname{Re} s_{z=\lambda}\left(z^{-s} \frac{1}{1 - \lambda^{-1} z} \frac{\mathrm{d}z}{z} \right) = \lambda^{-s}. \tag{7}$$

□

Example 2.4. $\mathbf{a} = (1/\nu!)_{\nu \geq 0}$ defines a class in $\mathcal{A}$ and $I(\mathbf{a}) = \Gamma(s+1)^{-1} \in \mathcal{F}$.

Proof. $\psi_{\mathbf{a}}(x) = \mathrm{e}^{x}$ is clearly good and

$$I(\psi_{\mathbf{a}})(s) = \frac{1}{2\pi i} \int_{-\infty}^{(0+)} z^{-s} \mathrm{e}^{z} \frac{\mathrm{d}z}{z} = \Gamma(s+1)^{-1}$$

is Hankel's representation of the inverse Γ-function. □

We can also argue as follows: Since $\psi_{\mathbf{a}}(x) = \mathrm{e}^{x}$ is good for any $\alpha \in \mathbb{R}$, lemma 1.3 shows that

$$I(\psi_{\mathbf{a}})(s) = -\frac{\sin \pi s}{\pi} M(\mathrm{e}^{-x})(-s) \quad \text{for } \operatorname{Re} s < 0.$$

Now by its definition $\Gamma(s)$ equals the Mellin transform of e^{-x} so that

$$I(\psi_{\mathbf{a}})(s) = -\frac{\sin \pi s}{\pi} \Gamma(-s) = \Gamma(s+1)^{-1}$$

first in $\operatorname{Re} s < 0$ and then for all s by analytic continuation.

Example 2.5. We want to interpolate the values $-B_{\nu+1}/(\nu+1)$ of the zeta-function at the negative integers. Since they grow so quickly that ψ has radius of convergence zero we renormalize them as follows: $-B_{\nu+1}/(\nu+1)!$ for $\nu \geq -1$. We expect them to be interpolated by the function $\zeta(-s)/\Gamma(s+1)$ and this is indeed the case. More generally consider the sequence: $(-B_{\nu+1}(a)/(\nu+1)!)_{\nu \geq -1}$ for $0 < a \leq 1$ where $B_n(a)$ is the nth Bernoulli polynomial. Its ψ-function is

$$\psi(z) = \sum_{\nu=-1}^{\infty} -B_{\nu+1}(a) \frac{z^{\nu}}{(\nu+1)!} = -\frac{1}{z} \sum_{\nu=0}^{\infty} B_{\nu}(a) \frac{z^{\nu}}{\nu!} = \frac{\mathrm{e}^{az}}{1 - \mathrm{e}^{z}}. \tag{8}$$

It is good and $|\psi(z)| = O(|z|^{\alpha})$ for any $\alpha \in \mathbb{R}$. We have

$$I(\psi)(s) = \frac{1}{2\pi i} \int_{-\infty}^{(0+)} z^{-s} \frac{\mathrm{e}^{az}}{1 - \mathrm{e}^{z}} \frac{\mathrm{d}z}{z} = \frac{\zeta(-s, a)}{\Gamma(s+1)} \tag{9}$$

by a standard formula from the theory of the Hurwitz zeta function $\zeta(s, a) = \sum_{\nu=0}^{\infty} (\nu + a)^{-s}$ c.f. [EMOT] 1.10. Thus $\frac{\zeta(-s,a)}{\Gamma(s+1)} \in \mathcal{F}$ is the interpolation of its values $\left(-\frac{B_{\nu+1}(a)}{(\nu+1)!}\right)_{\nu \geq \nu_0}$ for any $\nu_0 \geq -1$. It follows that $\frac{L(\chi, -s)}{\Gamma(s+1)} \in \mathcal{F}$ is the interpolation of its values at the integers $\nu \geq \nu_0$ for any $\nu_0 \geq -1$ as well.

4. Invoking the Hardy, Ramanujan theory. Further examples

The problem is of course to give good criteria as to when an analytic function defined in some right half plane belongs to $\mathcal{F}$. For this we take up ideas of Hardy.

We first require a formula of Hardy, [H], (11.4.4) whose proof is omitted in [H]. For the convenience of the reader we give a proof below. Actually, in the following proposition, we show a slightly stronger result since this requires no extra effort and may be useful for extending the theory.

PROPOSITION 3.1

Assume that ϕ is holomorphic in $\operatorname{Re} s > \beta$ *and satisfies an estimate of the form:*

$$|\phi(\sigma + it)| \leq f(t)\mathrm{e}^{P\sigma+\pi|t|} \quad \textit{for } \sigma > \beta$$

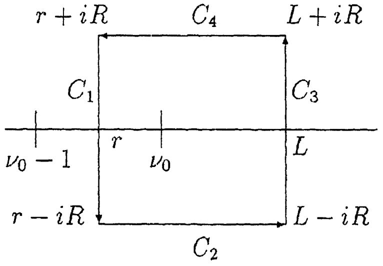
where $P \in \mathbb{R}$ and $f \in L^1(\mathbb{R})$ is such that $\lim_{t\to\pm\infty} f(t) = 0$.

Fix an integer $\nu_0 > \beta$ and choose $r > \beta$ such that $\nu_0 - 1 < r < \nu_0$. Then for any real $-\mathrm{e}^{-P} < x < 0$ we have the integral representation:

$$\psi(x) = \sum_{\nu \geq \nu_0} \phi(\nu)x^\nu = -\frac{1}{2\pi i}\int_{r-i\infty}^{r+i\infty} \frac{\pi}{\sin \pi s}\phi(s)(-x)^s \mathrm{d}s.$$

Here the series is absolutely convergent and the integral is in the Lebesgue sense.

Proof. Consider the contour $C = C_1 + C_2 + C_3 + C_4$:



where $L \in \frac{1}{2} + \mathbb{Z}$. By the residue theorem:

$$\sum_{\nu_0 \leq \nu < L} \phi(\nu)x^\nu = \frac{1}{2\pi i}\int_C \frac{\pi}{\sin \pi s}\phi(s)(-x)^s \, \mathrm{d}x.$$

We have

$$\left|\frac{\pi}{\sin \pi s}\right| \ll \mathrm{e}^{-\pi|\operatorname{Im} s|} \quad \text{for } |\operatorname{Im} s| \gg 0.$$

Using periodicity of sin we get that for R large enough

$$\left|\frac{\pi}{\sin \pi s}\right| \leq c_1\mathrm{e}^{-\pi|\operatorname{Im} s|} \text{ holds on } C \text{ for all } L.$$

Hence

$$\left|\int_{C_2}\right| \leq c_1(L - r)\mathrm{e}^{-\pi R}\mathrm{e}^{\max(Pr,PL)+\pi R}\max((-x)^r, (-x)^L)f(R).$$

Thus for fixed L, we have $\lim_{R\to\infty}\int_{C_2} = 0$. Similarly $\lim_{R\to\infty}\int_{C_4} = 0$.
Next

$$\left|\int_{L-i\infty}^{L+i\infty}\right| \leq c_2 \int_{-\infty}^{\infty} \mathrm{e}^{-\pi|t|}\mathrm{e}^{PL}\mathrm{e}^{\pi|t|}f(t)(-x)^L\,\mathrm{d}t \leq c_3\mathrm{e}^{L(P+\log(-x))}\int_{-\infty}^{\infty} f(t)\,\mathrm{d}t.$$

Hence the integral exists and tends to zero for $L\to\infty$ by our assumption $-\mathrm{e}^{-P} < x < 0$ i.e. $P+\log(-x) < 0$. Similarly the integral from $r-i\infty$ to $r+i\infty$ exists. Hence the formula. □

One now uses the integral representation for ψ of the proposition to show that ψ which *a priori* is holomorphic only in $0 < |z| < \mathrm{e}^{-P}$ extends to a holomorphic function in some punctured neighborhood U' as in (2) above which is bounded by a power of $|z|$ as in (3). More can be done but let us stay with a class of functions introduced by Hardy. For $A < \pi$ set:

$$\mathcal{F}_H(A) = \left\{ \begin{array}{l} \phi\text{'s analytic in } \operatorname{Re} s > \beta \text{ for some } \beta \in \mathbb{R} \text{ such that there} \\ \text{exists } P \in \mathbb{R} \text{ with } |\phi(\sigma+it)| \ll \mathrm{e}^{P\sigma + A|t|} \text{ in } \operatorname{Re} s > \beta \end{array} \right\}.$$

Any such β is called allowable for ϕ. Set $\mathcal{F}_H = \bigcup_{A<\pi}\mathcal{F}_H(A)$. Then we have the following result which follows from the preceeding considerations and those in [H], 11.4:

Theorem 3.2. *$\mathcal{F}_H \subset \mathcal{F}$. More precisely, if β is allowable for $\phi \in \mathcal{F}_H$ let ψ be the analytic continuation of $\sum_{\nu>\beta}\phi(\nu)x^\nu$ to a punctured neighborhood U' of $(-\infty,0]$. Then we have $|\psi(z)| = O(|z|^\alpha)$ as $|z|\to\infty$ in U' for every $\alpha > \beta$ and the interpolation formula*

$$I(\psi)(s) = \frac{1}{2\pi i}\int_{-\infty}^{(0+)} z^{-s}\psi(z)\frac{\mathrm{d}z}{z} = \phi(s)$$

therefore holds in $\operatorname{Re} s > \beta$.

Remark. The example of $\phi(s) = \sin \pi s$ shows that the condition $A < \pi$ is not unnatural.

Proof. By assumption $|\phi(\sigma+it)| \ll \mathrm{e}^{P\sigma+A|t|}$ in $\sigma > \beta$ for some $P \in \mathbb{R}, A < \pi$. Hence Proposition 3.1 is applicable. Let ν_0 be the least integer $> \beta$ and choose $r > \beta$ such that $\nu_0 - 1 < r < \nu_0$. Then by (3.1) we have for any $-\mathrm{e}^{-P} < z < 0$:

$$\psi(z) = -\frac{1}{2\pi i}\int_{r-i\infty}^{r+i\infty} \frac{\pi}{\sin \pi s}\phi(s)(-z)^s\,\mathrm{d}s.$$

Choose $0 < \delta < \pi - A$. Then for $-\delta < \arg(-z) < \delta$ we have:

$$|(-z)^s| = |z|^\sigma \mathrm{e}^{-t\arg(-z)} \leq |z|^\sigma \mathrm{e}^{\delta|t|}.$$

Thus

$$\begin{aligned}\left|\int_{r-i\infty}^{r+i\infty}\right| &\leq c_1\int_{-\infty}^{\infty} \mathrm{e}^{-\pi|t|}\mathrm{e}^{Pr+A|t|}|z|^r\mathrm{e}^{\delta|t|}\,\mathrm{d}t \\ &\leq c_2|z|^r\int_{-\infty}^{\infty}\mathrm{e}^{(A+\delta-\pi)|t|}\,\mathrm{d}t = O(|z|^r).\end{aligned}$$

Since we know that the series for ψ converges in $0 < |z| < e^{-P}$ it follows that ψ extends to an analytic function in some region U' as in (2) where it satisfies $\psi(z) = O(|z|^r)$ as $|z| \to \infty$ for any $r > \beta$. Thus ψ is good and hence $\phi \in \mathcal{F}'$. Moreover $I(\psi)(s)$ defines a holomorphic function in $\operatorname{Re} s > \beta$. It remains to prove that $I(\psi) = \phi$. Unfortunately this cannot be checked by substituting the above integral representation for ψ into the contour integral I since the former does not converge for the s on the loop around zero. Instead we reduce the claim to a formula of Hardy and Ramanujan – the last equality in [H], 11.4– which itself is an application of Mellin- or Fourier-inversion:

Formula of Hardy–Ramanujan. For $0 < \delta < 1$, let ϕ_{H} be holomorphic in $\operatorname{Re} s \geq -\delta$ and satisfy the estimate $\phi_{\mathrm{H}}(s) \ll e^{P_1 \sigma + A|t|}$ there for some P_1 and $A < \pi$. Setting

$$\Phi_{\mathrm{H}}(x) = \sum_{\nu=0}^{\infty} \phi_{\mathrm{H}}(\nu)(-1)^{\nu} x^{\nu}$$

we have that

$$\int_0^{\infty} x^w \Phi_{\mathrm{H}}(x) \frac{dx}{x} = \frac{\pi}{\sin \pi w} \phi_{\mathrm{H}}(-w) \quad \text{for } 0 < \operatorname{Re} w < \delta.$$

By Lemma 1.3 we have for $\beta < \operatorname{Re} s < \nu_0$:

$$I(\psi)(s) = -\frac{\sin \pi s}{\pi} \int_0^{\infty} x^{-s} \psi(-x) \frac{dx}{x}.$$

Now choose $0 < \delta < \nu_0 - \beta$ so that in particular $\delta < 1$. Set $\phi_{\mathrm{H}}(s) = \phi(s + \nu_0)$. Then the Hardy–Ramanujan formula applied to $\phi_{\mathrm{H}}(s) = \phi(s + \nu_0)$ gives the equality:

$$\int_0^{\infty} x^{w-\nu_0} \psi(-x) \frac{dx}{x} = \frac{\pi}{\sin \pi(w - \nu_0)} \phi(\nu_0 - w).$$

Thus for $\nu_0 - \delta < \operatorname{Re} s < \nu_0$ we find that

$$-\frac{\sin \pi s}{\pi} \int_0^{\infty} x^{-s} \psi(-x) \frac{dx}{x} = \phi(s).$$

Together with the above formula for $I(\psi)(s)$ it follows by analytic continuation that

$$I(\psi)(s) = \phi(s)$$

for $\operatorname{Re} s > \beta$ as claimed. □

Remark 3.3. Our interpolation functional I has two advantages over the one of Hardy–Ramanujan:

$$I_{\mathrm{HR}} : \psi \longmapsto \frac{\sin \pi s}{\pi} \int_0^{-\infty} (-x)^{-s} \psi(x) \frac{dx}{x}$$

which requires convergence at 0 and $-\infty$ whereas I needs convergence at $-\infty$ only. As a consequence interpolation formulas involving I_{HR} are valid at most in some region $\beta < \operatorname{Re} s < \gamma$ whereas those using I hold in a half plane $\operatorname{Re} s > \beta$. Moreover only in I is it possible to add to ψ an arbitrary Laurent polynomial without changing its value. This is crucial for interpolating elements of $\mathcal{A}$ i.e. sequences which are only given up to equivalence.

In the rest of this section we use distributions to give a different and more conceptual proof of the assertion $\phi \in \mathcal{F}$ in Theorem 3.2 for a restricted class of functions ϕ:

Let ϕ be an entire function which satisfies an estimate of the form

$$|\phi(s)| \ll (1+|s|)^N \mathrm{e}^{A|\operatorname{Im} s|}$$

in $\mathbb{C}$ for some $A < \pi$. By Schwartz' extension of the Paley–Wiener theorem to distributions [Y], VI.4 the function ϕ is the Fourier–Laplace transform of a distribution T with compact support in $(-\pi, \pi)$. Choose some $\varepsilon > 0$ such that $\operatorname{supp} T$ is disjoint from the set C_ε of y in $\mathbb{R}$ with $|\mathrm{e}^{iy} + 1| < \varepsilon$. Let α be a smooth function on $\mathbb{R}$ which is 0 on C_ε and equal to 1 on $\operatorname{supp} T$. We have:

$$\phi(s) = \hat{T}(s) = (2\pi)^{-1/2}\langle T_y, \mathrm{e}^{-isy}\rangle.$$

Hence

$$\psi(x) := \sum_{\nu=0}^{\infty} \phi(\nu)x^\nu = (2\pi)^{-1/2}\langle T_y, (1 - x\mathrm{e}^{-iy})^{-1}\rangle$$

for $|x| < 1$. The formula

$$\psi(z) = (2\pi)^{-1/2}\langle T_y, \alpha(y)(1 - z\mathrm{e}^{-iy})^{-1}\rangle$$

gives the analytic continuation of ψ to a neighborhood of $(-\infty, 0]$. Since

$$|T(h)| \le C \sum_{|\beta| \le N} \sup_{|y| \le L} |D^\beta h(y)|$$

for some constants C, N, L and all smooth functions h on $\mathbb{R}$ it follows that

$$|\psi(z)| = O(|z|^{-1}) \quad \text{as } z \to -\infty.$$

Hence ψ is good and thus $\phi \in \mathcal{F}'$. Now

$$\begin{aligned} I(\psi)(s) &= \frac{1}{2\pi i}\int_{-\infty}^{(0+)} z^{-s}(2\pi)^{-1/2}\langle T_y, \alpha(y)(1 - z\mathrm{e}^{-iy})^{-1}\rangle \frac{\mathrm{d}z}{z} \\ &= (2\pi)^{-1/2}\left\langle T_y, \frac{1}{2\pi i}\int_{-\infty}^{(0+)} z^{-s}\alpha(y)(1 - z\mathrm{e}^{-iy})^{-1}\frac{\mathrm{d}z}{z}\right\rangle \\ &\overset{2.3}{=} (2\pi)^{-1/2}\langle T_y, \mathrm{e}^{-isy}\rangle = \phi(s). \end{aligned}$$

Since $\operatorname{supp} T \subset (-\pi, \pi)$. Hence $\phi \in \mathcal{F}$.

If more generally T has compact support in $\mathbb{R} \setminus \pi\mathbb{Z}$ then ψ is still good by the identical argument, so that $\phi \in \mathcal{F}'$. However we now have, again using (2.3), that:

$$P\phi = I(\psi)(s) = (2\pi)^{-1/2}\langle T_y, \alpha(y)\mathrm{e}^{-is\bar{y}}\rangle, \tag{10}$$

where $\bar{y} \in (-\pi, \pi]$ is such that $\bar{y} \equiv y \bmod \pi\mathbb{Z}$. Note that $\mathrm{e}^{-is\bar{y}}$ is not smooth but $\alpha(y)\mathrm{e}^{-is\bar{y}}$ is. Writing T as a finite sum

$$T = \sum_\nu T_\nu$$

of distributions T_ν with compact support in $(-\pi + \nu\pi, \pi + \nu\pi)$ it follows that $\phi = \sum_\nu \phi_\nu$ where $\phi = \hat{T}, \phi_\nu = \hat{T}_\nu$. By (10) we see that $P(\phi_\nu) = e^{i\nu\pi s}\phi_\nu$ and hence

$$P\phi = \sum_\nu e^{i\nu\pi s}\phi_\nu \in \mathcal{F}.$$

Incidentially this is also a consequence of Theorem 3.2 applied to $e^{i\nu\pi s}\phi_\nu$.

5. Applications

In this section we illustrate the preceeding theory by interpolating certain interesting classes of functions. We are mostly interested in L-series and their completed versions by Γ-factors.

Set

$$\mathcal{F}_{\mathrm{H}}^{0+} = \left\{ \begin{array}{c} \phi\text{'s analytic in } \mathrm{Re}\, s > \beta \text{ for some } \beta \in \mathbb{R} \text{ s.t. for every} \\ \delta > 0 \text{ there exist a } \beta_\delta \geq \beta \text{ and some } P_\delta \in \mathbb{R} \text{ with} \\ |\phi(\sigma + it)| \ll e^{P_\delta \cdot \sigma + \delta \cdot |t|} \text{ in } \mathrm{Re}\, s > \beta_\delta \end{array} \right\}$$

and

$$\mathcal{F}_{\mathrm{H}}^{0} = \left\{ \begin{array}{l} \phi\text{'s analytic in } \mathrm{Re}\, s > \beta \text{ for some `associated' } \beta \text{ s.t.} \\ |\phi(\sigma + it)| \ll e^{P \cdot \sigma} \text{ in } \mathrm{Re}\, s > \beta \text{ for some } P \in \mathbb{R} \end{array} \right\}.$$

Clearly $\mathcal{F}_{\mathrm{H}}^0 \subset \mathcal{F}_{\mathrm{H}}^{0+} \subset \mathcal{F}_{\mathrm{H}}$. Moreover $\mathcal{F}_{\mathrm{H}}^0$ and $\mathcal{F}_{\mathrm{H}}^{0+}$ are $\mathbb{C}$-algebras and $\mathcal{F}_{\mathrm{H}}$ is a module under them. Note that if $f \in \mathcal{F}_{\mathrm{H}}^0$ and $\phi \in \mathcal{F}_{\mathrm{H}}$ have β_f and β_ϕ associated to them, then $\max(\beta_f, \beta_\phi)$ is associated to $f\phi$.

Clearly every Dirichlet series $\sum a_n \lambda_n^{-s}$ with $\lambda_n > 0$ and abscissa of absolute convergence $\beta < \infty$ belongs to $\mathcal{F}_{\mathrm{H}}^0$ with β being admissible. In particular L-series and their inverses belong to $\mathcal{F}_{\mathrm{H}}^0$.

On the other hand L-series completed by Γ-factors do not even belong to $\mathcal{F}'$ since the associated power series ψ has radius of convergence zero. The reciprocal function however has a better behaviour if the Γ-factor is simple. To see this we require the following fact:

PROPOSITION 4.1

For every $0 < a < 2$, $b \in \mathbb{R}$ *the function* $\Gamma(as + b)^{-1}$ *belongs to* $\mathcal{F}_{\mathrm{H}}$ *with associated* $\beta = \frac{1}{a}(\frac{1}{2} - b)$.

Proof. For given $\delta > 0$ the complex Stirling asymptotics for $\Gamma(s)$ implies that

$$\Gamma(s) = e^{-s} e^{(\sigma - 1/2)\log s} (2\pi)^{1/2} (1 + O(s^{-1}))$$

in $|\arg s| \leq \pi - \delta$ as $|s| \to \infty$. Hence this estimate holds for all s with $\mathrm{Re}\, s \geq \frac{1}{2}, |s| > 1$. Thus we also have

$$\Gamma(s)^{-1} = e^{s} e^{(1/2 - s)\log s} (2\pi)^{-1/2} (1 + O(s^{-1})) \quad \text{in } \mathrm{Re}\, s \geq \tfrac{1}{2}, |s| > 1$$

and hence:

$$\begin{aligned} |\Gamma(s)^{-1}| &\ll e^{\sigma} e^{(1/2 - \sigma)\log|s|} e^{|t|\pi/2} \\ &\ll e^{\sigma + |t|\pi/2} \end{aligned}$$

in $\operatorname{Re} s \geq 1/2, |s| > 1$ and hence in $\operatorname{Re} s \geq 1/2$. Thus

$$|\Gamma(as+b)^{-1}| \ll e^{a\sigma+|t|a/2\pi} \quad \text{for } \operatorname{Re} s \geq \tfrac{1}{a}(\tfrac{1}{2}-b). \tag{11}$$

□

Examples. (1) It follows again that $\Gamma(s)^{-1} \in \mathcal{F}$.
(2) Since $\zeta(2s+2)^{-1} \in \mathcal{F}_{\mathrm{H}}^{0}$ with $\beta = -1/2$ (abscissa of absolute convergence) and since $\Gamma(s+1)^{-1} \in \mathcal{F}_{\mathrm{H}}$ with $\beta = -1/2$ by the proposition we find that $\Gamma(s+1)^{-1}\zeta(2s+2)^{-1} \in \mathcal{F}_{\mathrm{H}}$ with $\beta = -1/2$. Hence theorem 3.2 gives us:

$$\frac{1}{2\pi i}\int_{-\infty}^{(0+)} z^{-s}\sum_{\nu=0}^{\infty}\frac{z^{\nu}}{\nu!\zeta(2\nu+2)}\frac{dz}{z} = \frac{1}{\Gamma(s+1)\zeta(2s+2)} \quad \text{for } \operatorname{Re} s > -\frac{1}{2}.$$

Note that the series in the integral converges everywhere. Setting

$$\hat{\zeta}(s) = \pi^{-s/2}\Gamma\left(\frac{s}{2}\right)\zeta(s)$$

we get similarly that $\hat{\zeta}(2s+2)^{-1} \in \mathcal{F}_{\mathrm{H}}$ with $\beta = -1/2$ and that:

$$\frac{1}{2\pi i}\int_{-\infty}^{(0+)} z^{-s}\sum_{\nu=0}^{\infty}\frac{z^{\nu}}{\hat{\zeta}(2\nu+2)}\frac{dz}{z} = \frac{1}{\hat{\zeta}(2s+2)} \quad \text{in } \operatorname{Re} s > -\frac{1}{2}.$$

Similarly $\hat{\zeta}(s)^{-1} \in \mathcal{F}_{\mathrm{H}}$ with $\beta = 1$ and hence:

$$\frac{1}{2\pi i}\int_{-\infty}^{(0+)} z^{-s}\sum_{\nu=2}^{\infty}\frac{z^{\nu}}{\hat{\zeta}(\nu)}\frac{dz}{z} = \frac{1}{\hat{\zeta}(s)} \quad \text{in } \operatorname{Re} s > 1.$$

(3) A similar formula holds for the completed L-series

$$\hat{L}(E,s) = L(E,s)(2\pi)^{-s}\Gamma(s)$$

of an elliptic curve E over $\mathbb{Q}$:

$$\frac{1}{2\pi i}\int_{-\infty}^{(0+)} z^{-s}\sum_{\nu=2}^{\infty}\frac{z^{\nu}}{\hat{L}(E,\nu)}\frac{dz}{z} = \frac{1}{\hat{L}(E,s)} \quad \text{in } \operatorname{Re} s > \frac{3}{2}.$$

(4) For $\phi(s) = \zeta(2s) \in \mathcal{F}_{\mathrm{H}}^{0} \subset \mathcal{F}$ and $\beta = 1/2$ the corresponding function ψ in Theorem 3.2 is given by $\psi(z) = f(\sqrt{-z})$ where f is the even function

$$f(w) = \frac{1}{2}\left[\frac{(1-\pi w)e^{\pi w} - (1+\pi w)e^{-\pi w}}{e^{\pi w} - e^{-\pi w}}\right].$$

After some calculation which we leave to the reader the formula of theorem 3.2 leads to the functional equation of $\zeta(s)$. This example was suggested by the discussion of Ramanujan's formula in [E], 10.10.

Remark. A variant of the first formula was first given by Riesz as mentioned by Hardy:

$$\int_{0}^{\infty} x^{-s}\sum_{\nu=0}^{\infty}\frac{(-1)^{\nu}x^{\nu}}{\nu!\zeta(2\nu+2)}\frac{dx}{x} = \frac{\Gamma(s)}{\zeta(2s+2)}$$

valid for $-1/2 < \operatorname{Re} s < 0$.

The case $\Gamma(2s+b)^{-1}$ is not covered by the proposition. We close by noting that a direct computation gives:

Fact 4.2. $\Gamma(2s+n)^{-1} \in \mathcal{F}$ for all $n \in \mathbb{Z}$.

Proof. Since $\mathcal{F}$ is shift-invariant we may restrict to $\Gamma(2s+1)^{-1}$. The associated function ψ is $\psi(x) = \sum_{\nu=0}^{\infty} x^{\nu}/(2\nu)!$ which is entire. We have $\psi(w^2) = \frac{1}{2}(\mathrm{e}^{w} + \mathrm{e}^{-w})$. The mapping $w \mapsto w^2$ transforms any strip $0 \leq \operatorname{Re} w \leq \delta$ into a neighborhood U of $(-\infty, 0]$. In $0 \leq \operatorname{Re} w \leq \delta$ the function $\frac{1}{2}(\mathrm{e}^{w} + \mathrm{e}^{-w})$ is bounded and hence ψ is bounded in U. Thus $\Gamma(2s+1)^{-1} \in \mathcal{F}'$. For $\operatorname{Re} s > 0$ we have:

$$\begin{aligned}\frac{1}{2\pi i}\int_{-\infty}^{(0+)} z^{-s}\psi(z)\frac{\mathrm{d}z}{z} &= \frac{1}{2\pi i}\int_{-\infty}^{(0+)} z^{-s}(\psi(z)-1)\frac{\mathrm{d}z}{z} - \frac{1}{2\pi i s}[z^{-s}]_{-\infty-0\cdot i}^{-\infty+0\cdot i}\\ &= \frac{1}{2\pi i}\int_{-\infty}^{(0+)} z^{-s}(\psi(z)-1)\frac{\mathrm{d}z}{z}.\end{aligned}$$

Using Lemma (1.3) we see that for $0 < \operatorname{Re} s < 1$ this equals

$$\begin{aligned}-\frac{\sin \pi s}{\pi}\int_{0}^{\infty} x^{-s}(\psi(-x)-1)\frac{\mathrm{d}x}{x} &= -2\frac{\sin \pi s}{\pi}\int_{0}^{\infty} x^{-2s}(\cos x - 1)\frac{\mathrm{d}x}{x}\\ &= \frac{\sin \pi s}{\pi s}\int_{0}^{\infty} x^{-2s}\sin x \,\mathrm{d}x\end{aligned}$$

by integration by parts. Substituting the formula in [EMOT], 1.5.1 (38)

$$\int_{0}^{\infty} x^{\alpha-1}\sin x\,\mathrm{d}x = \Gamma(\alpha)\sin\frac{\pi}{2}\alpha \quad \text{in } -1 < \operatorname{Re}\alpha < 1.$$

We arrive after some calculation at the desired formula:

$$\frac{1}{2\pi i}\int_{-\infty}^{(0+)} z^{-s}\psi(z)\frac{\mathrm{d}z}{z} = \Gamma(2s+1)^{-1}.$$

□

Acknowledgements

I would like to thank B Mazur for his question and the Harvard mathematics department for its hospitality. I would also like to thank the referee for suggestions to improve the exposition.

References

[BK] Bloch S and Kato K, *L*-functions and Tamagawa numbers of motives, in: *The Grothendieck Festschrift, vol. 1, Prog. Math.* **86** (1990) 333–400

[E] Edwards H M, *Riemann's zeta function* (Academic Press) (1974)

[EMOT] Erdélyi A *et al*, *Higher transcendental functions.* The Bateman Manuscript Project (McGraw-Hill) (1953) vol. 1

[H] Hardy G H and Ramanujan S, *Twelve Lectures on Subjects Suggested by His Life and Work* (Chelsea) (1978)

[I] Igusa J-I, Lectures on forms of higher degree. (Bombay: Tata Institute of Fundamental research) (1978)

[Y] Yosida K, Grundlehren Bd. 123, (Springer: Functional Analysis) (1971)

Proc. Indian Acad. Sci. (Math. Sci.), Vol. 110, No. 2, May 2000, pp. 133–136.
© Printed in India

RRF rings which are not LRF

K VARADARAJAN

Department of Mathematics and Statistics, University of Calgary, Calgary, Alberta
T2N 1N4, Canada
E-mail: varadara@math.ucalgary.ca

MS received 9 August 1999

Abstract. Define a ring A to be RRF (respectively LRF) if every right (respectively left) A-module is residually finite. We determine the necessary and sufficient conditions for a formal triangular matrix ring $T = \begin{pmatrix} A & 0 \\ M & B \end{pmatrix}$ to be RRF (respectively LRF). Using this we give examples of RRF rings which are not LRF.

Keywords. Residual finiteness; simple modules; injective hulls; boolean rings.

1. Introduction

All the rings we consider will be associative rings with an identity element $1 \neq 0$ and all the modules considered will be unital modules. In what follows A denotes a ring, mod-A (respectively A-mod) will denote the category of right (respectively left) A-modules. Recall (Definition 2.1 in [9]) that an A-module M is said to be residually finite if given any $x \neq 0$ in M, we can find a submodule N of M (depending on x) with $x \notin N$ and M/N finite. Note that we require M/N to be actually finite.

DEFINITION 1.1

A is called an RRF (respectively LRF) ring if every right (respectively left) A-module is residually finite.

A will be called an RF ring if it is both RRF and LRF.

One of the questions raised in [10] is: Are there RRF rings which are not LRF? (open problem 2 in § 5 of [10]). As is well-known triangular matrix rings act as a good source for constructing many counter examples. For instance Herstein [4] constructed a counter example to a conjecture of Jacobson using triangular matrix rings. In [1] Goodearl used triangular matrix rings to construct examples of rings R over which all finitely generated left (as well as right) R-modules are cohopfian, all finitely generated left R-modules are hopfian but there exist cyclic non-hopfian right R-modules. Motivated by these, in this note we first obtain necessary and sufficient conditions for the formal triangular matrix ring $T = \begin{pmatrix} A & 0 \\ M & B \end{pmatrix}$ (where ${}_BM_A$ is a left B, right A bimodule) to be an RRF (respectively an LRF) ring. This then allows us to construct easily examples of RRF rings which are not LRF.

2. Necessary and sufficient conditions for a triangular matrix ring to be RRF (respectively LRF)

Let A, B be rings and ${}_BM_A$ be a left B, right A bimodule and $T = \begin{pmatrix} A & 0 \\ M & B \end{pmatrix}$ the formal triangular matrix ring whose elements are formal matrices $\begin{pmatrix} a & 0 \\ m & b \end{pmatrix}$ with $a \in A$, $b \in B$ and $m \in M$. Addition in T is defined co-ordinate wise and multiplication is given by $\begin{pmatrix} a & 0 \\ m & b \end{pmatrix}\begin{pmatrix} a' & 0 \\ m' & b' \end{pmatrix} = \begin{pmatrix} aa' & 0 \\ ma' + bm' & bb' \end{pmatrix}$. To determine the necessary and sufficient conditions for T to be RRF we will use Green's representation of mod-T [2] and the following proposition characterizing RRF rings proved in [10]. Actually combining this with recent work of Hirano [5, 6] we obtained structure theorems on RRF rings in [10]

PROPOSITION 2.1

The following conditions are equivalent for a ring A.

(1) *A is RRF.*
(2) *For every simple right A-module S the injective hull $E(S)$ of S is finite.*
(3) *There exists a cogenerator U for* mod-A *with U residually finite.*

This is actually Proposition 2.1 in [10].

We now give a brief description of Green's representation of mod-T. Let Ω denote the category whose objects are triples $(X, Y)_f$ where $X \in$ mod-A, $Y \in$ mod-B and $f : Y \otimes_B M \longrightarrow X$ is a map in mod-A. A morphism from $(X, Y)_f$ to $(U, V)_g$ in Ω consists of a pair (φ_1, φ_2) where $\varphi_1 : X \longrightarrow U$ is a map in mod-A and $\varphi_2 : Y \longrightarrow V$ is a map in mod-B satisfying the condition $\varphi_1 \circ f = g \circ (\varphi_2 \otimes \mathrm{Id}_M)$. It is known that Ω is equivalent to mod-T. The right T-module corresponding to the triple $(X, Y)_f$ under the equivalence is additively the direct sum $X \oplus Y$ with right T-action given by $(x, y)\begin{pmatrix} a & 0 \\ m & b \end{pmatrix} = (xa + f(y \otimes m), yb)$ for all $a \in A$, $b \in B$ and $m \in M$.

We need the description of the triple in Ω which corrresponds to the injective hull of the right T-module corresponding to $(X, Y)_f$. For any ring R and any R-module W the injective hull of W will be denoted by $E(W)$. Associated to $f : Y \otimes_B M \longrightarrow X$ we have the map $\widetilde{f} : Y \longrightarrow \mathrm{Hom}(M_A, X_A)$ given by $(\widetilde{f}(y))(m) = f(y \otimes m)$. The left B-action on M induces a right B-action on $\mathrm{Hom}(M_A, X_A)$. It is clear that $\widetilde{f} : Y \longrightarrow \mathrm{Hom}(M_A, X_A)$ is a map in mod-B. The following is due to Stenstrom [8].

PROPOSITION 2.2

The triple which corresponds to the injective hull of the right T-module determined by $(X, Y)_f$ is given by $(E(X_A), \mathrm{Hom}(M_A, E(X_A)) \oplus E(\ker \widetilde{f}_B))_\delta$ where δ: $\{\mathrm{Hom}(M_A, E(X_A)) \oplus E(\ker \widetilde{f}_B)\} \otimes_B M \longrightarrow E(X_A)$ is the map $\delta(\theta, u) \otimes m = \theta(m)$ for any $\theta \in \mathrm{Hom}(M_A, E(X_A))$ and $u \in E(\ker \widetilde{f}_B)$.

A more general result is stated on page 1995 of [7]. The main result of this section is the following:

Theorem 2.3. *Let T be the formal triangular matrix ring $\begin{pmatrix} A & 0 \\ M & B \end{pmatrix}$ where M is a left B right A bimodule. Then*

(1) *T is RRF if any only if A and B are RRF and* $\mathrm{Hom}(M_A, E(S_A))$ *is finite for every simple right A-module S_A.*

(2) *T is LRF if and only if A and B are LRF and* $\mathrm{Hom}({}_BM, E({}_BL))$ *is finite for every simple left B-module ${}_BL$.*

Proof. In [3] we determined all the triples in Ω corresponding to simple submodules of the right T-module determined by $(X, Y)_f$ in Ω (Proposition 2.1 in [3]). The same arguments show that the simple right T-modules correspond to triples of the form $(S_A, 0)_0$ or $(0, L_B)_0$ where S_A is simple in mod-A and L_B is simple in mod-B.

From Proposition 2.2 the respective injective hulls in mod-T correspond to the triples $(E(S_A), \mathrm{Hom}(M_A, E(S_A)))_\delta$ and $(0, E(L_B))_0$, where $\delta : \mathrm{Hom}(M_A, E(S_A)) \otimes_B M \to E(S_A)$ is given by $\delta(\varphi \otimes m) = \varphi(m)$ for any $\varphi \in \mathrm{Hom}(M_A, E(S_A))$. It follows that T is RRF$\Longleftrightarrow E(S_A)$, $\mathrm{Hom}(M_A, E(S_A))$ and $E(L_B)$ are all finite for any simple right A-module S_A and any simple right B-module L_B. Equivalently T is RRF$\Longleftrightarrow A$ is RRF, B is RRF and $\mathrm{Hom}(M_A, E(S_A))$ is finite for every simple right A-module S_A. This proves (1) of Theorem 2.3.

The category Ω' of triples $({}_AX, {}_BY)_g$ where $g : M \otimes_A X \to Y$ is a map in B-mod with the obvious definition of morphisms is equivalent to the category T-mod. The proof of (2) is similar to that of (1) using the category equivalence between Ω' and T-mod. We omit the proof.

3. RRF rings which are not LRF

As seen already in [10] any boolean ring is RF. Let A be any infinite boolean ring. Then A is an infinite dimensional vector space over $\mathbb{Z}/2\mathbb{Z}$. We regard A as a left $\mathbb{Z}/2\mathbb{Z}$ and right A bimodule and consider the formal triangular matrix ring $T = \begin{pmatrix} A & 0 \\ A & \mathbb{Z}/2\mathbb{Z} \end{pmatrix}$. For every maximal ideal I of A, we have $A/I \simeq \mathbb{Z}/2\mathbb{Z}$. A being a commutative regular ring, it is a V-ring. Hence $E(A/I) = A/I \simeq \mathbb{Z}/2\mathbb{Z}$. From $\mathrm{Hom}(A_A, E(A/I)) \simeq E(A/I) \simeq \mathbb{Z}/2\mathbb{Z}$ we see that $\mathrm{Hom}(A_A, E(A/I))$ is finite. From (1) of Theorem 2.3 we conclude that T is RRF. Since A is an infinite dimensional vector space over $\mathbb{Z}/2\mathbb{Z}$ we see that $\mathrm{Hom}({}_{\mathbb{Z}/2\mathbb{Z}}A, \mathbb{Z}/2\mathbb{Z})$ is infinite. From (2) of Theorem 2.3 we see that T is not LRF. Thus we have proved the following.

Theorem 3.1. *For any infinite boolean ring A, the formal triangular matrix ring* $T = \begin{pmatrix} A & 0 \\ A & \mathbb{Z}/2\mathbb{Z} \end{pmatrix}$ *is an RRF ring which is not LRF.*

References

[1] Goodearl K R, Surjective endomorphisms of finitely generated modules, *Comm. Algebra* **15** (1987) 589–609

[2] Green E L, On the representation theory of rings in matrix form, *Pacific J. Math.* **100** (1982) 123–138

[3] Haghany A and Varadarajan K, Study of modules over formal triangular matrix rings, to appear in *J. Pure Appl. Algebra* (2000)

[4] Herstein I N, A counterexample in Noetherian rings, *Proc. Natl. Acad. Sci. U.S.A.* **54** (1965) 1036–1037

[5] Hirano Y, On rings over which each module has a maximal submodule, Preprint

[6] Hirano Y, Rings in which the injective hulls of simple modules have finiteness conditions, Preprint

[7] Muller Marianne, Rings of quotients of generalized matrix rings, *Comm. Algebra* **15** (1987) 1991–2015

[8] Stenstrom B, The maximal ring of quotients of a generalised matrix ring, Universale Algebren und theorie der radikale, Studien zur Algebra und ihre Anwendungen (Akademie-Verlag, Berlin) (1976) pp. 65–67

[9] Varadarajan K, Residual finiteness in rings and modules, *J. Ramanujan Math. Soc.* **8** (1993) 29–48

[10] Varadarajan K, Rings with all modules residually finite, *Proc. Indian Acad. Sci.* (*Math. Sci.*) **109** (1999) 345–351

Proc. Indian Acad. Sci. (Math. Sci.), Vol. 110, No. 2, May 2000, pp. 137–146.
© Printed in India

Inequalities for a polynomial and its derivative

V K JAIN

Mathematics Department, Indian Institute of Technology, Kharagpur 721 302, India

MS received 29 July 1999; revised 20 December 1999

Abstract. For an arbitrary entire function f and any $r > 0$, let $M(f, r) := \max_{|z|=r} |f(z)|$. It is known that if p is a polynomial of degree n having no zeros in the open unit disc, and $m := \min_{|z|=1} |p(z)|$, then

$$M(p', 1) \leq \frac{n}{2}\{M(p, 1) - m\},$$

$$M(p, R) \leq \left(\frac{R^n + 1}{2}\right)M(p, 1) - m\left(\frac{R^n - 1}{2}\right), \quad R > 1.$$

It is also known that if p has all its zeros in the closed unit disc, then

$$M(p', 1) \geq \frac{n}{2}\{M(p, 1) + m\}.$$

The present paper contains certain generalizations of these inequalities.

Keywords. Inequalities; zeros; polynomial.

1. Introduction and statement of results

Let $p(z)$ be a polynomial of degree n. Concerning the estimate of $|p'(z)|$ on the disc $|z| \leq 1$, we have the following famous result known as Bernstein's inequality [11].

Theorem A. *If $p(z)$ is a polynomial of degree n, then*

$$M(p', 1) \leq nM(p, 1), \tag{1.1}$$

with equality only for $p(z) = \alpha z^n$.

For polynomials having no zeros in $|z| < 1$, Erdös conjectured and Lax [5] proved

Theorem B. *If $p(z)$ is a polynomial of degree n, having no zeros in $|z| < 1$, then*

$$M(p', 1) \leq \frac{n}{2}M(p, 1), \tag{1.2}$$

with equality for those polynomials, which have all their zeros on $|z| = 1$.

For polynomials having all their zeros in $|z| \leq 1$, Turan [12] proved

Theorem C. *If $p(z)$ is a polynomial of degree n, having all its zeros in $|z| \leq 1$, then*

$$M(p', 1) \geq \frac{n}{2}M(p, 1), \tag{1.3}$$

with equality for those polynomials, which have all their zeros on $|z| = 1$.

On the other hand, concerning the estimate of $|p(z)|$ on the disc $|z| \leq R, R > 1$, w have, as a simple consequence of maximum modulus principle [7].

Theorem D. *If $p(z)$ is a polynomial of degree n, then*

$$M(p,R) \leq R^n M(p,1), \quad R > 1, \qquad (1.4$$

with equality for $p(z) = \alpha z^n$.

For polynomials not vanishing in $|z| < 1$, Ankeny and Rivlin [1] proved

Theorem E. *If $p(z)$ is a polynomial of degree n, having no zeros in $|z| < 1$, then*

$$M(p,R) \leq \frac{R^n + 1}{2} M(p,1), \quad R > 1, \qquad (1.5$$

with equality for $p(z) = \alpha + \beta z^n, |\alpha| = |\beta|$.

In [3], we had used a parameter β and obtained the following generalizations o inequalities (1.2), (1.5) and (1.3).

Theorem F. *Let $p(z)$ be a polynomial of degree n, having no zeros in $|z| < 1$. I $M(p,1) = 1$, then for $|\beta| \leq 1$*

$$\left|zp'(z) + \frac{n\beta}{2}p(z)\right| \leq \frac{n}{2}\left\{\frac{|\beta|}{2} + \left|1 + \frac{\beta}{2}\right|\right\}, \quad |z| = 1, \qquad (1.6$$

$$\left|p(Rz) + \beta\left(\frac{R+1}{2}\right)^n p(z)\right| \leq \frac{1}{2}\left\{\left|+\beta\left(\frac{R+1}{2}\right)^n\right| + \left|R^n + \beta\left(\frac{R+1}{2}\right)^n\right|\right\},$$
$$R \geq 1, \quad |z| = 1. \qquad (1.7$$

The result is best possible and equality holds in (1.6) and (1.7) for $p(z) = \alpha + \gamma z^n$, wit $|\alpha| = |\gamma|$.

Theorem G. *If $p(z)$ is a polynomial of degree n, having all its zeros in the closed uni disc, then for $|\beta| \leq 1$*

$$\max_{|z|=1}\left|zp'(z) + \frac{n\beta}{2}p(z)\right| \geq \frac{n}{2}\{1 + \mathrm{Re}(\beta)\}M(p,1). \qquad (1.8$$

Aziz and Dawood [2] used

$$m = \min_{|z|=1} |p(z)| \qquad (1.9$$

to obtain certain refinements of inequalities (1.2), (1.5) and (1.3) and proved

Theorem H. *If $p(z)$ is a polynomial of degree n which does not vanish in $|z| < 1$, the*

$$M(p',1) \leq \frac{n}{2}[M(p,1) - m], \qquad (1.10$$

$$M(p,R) \leq \left(\frac{R^n + 1}{2}\right)M(p,1) - \left(\frac{R^n - 1}{2}\right)m, \quad R > 1. \qquad (1.11$$

The result is best possible and equality holds in (1.10) and (1.11) for $p(z) = \alpha z^n +$ with $|\alpha| \leq |\gamma|$.

Theorem I. *If $p(z)$ is a polynomial of degree n which has all its zeros in $|z| \leq 1$, then*

$$M(p', 1) \geq \frac{n}{2}\{M(p, 1) + m\}. \tag{1.12}$$

The result is best possible and equality in (1.12) *holds for $p(z) = \alpha z^n + \gamma$, $|\gamma| \leq |\alpha|$.*

In this paper, we have used a parameter β, to obtain generalizations of inequalities (1.10), (1.11) and (1.12), similar to the generalizations – namely Theorems F and G, of inequalities (1.2), (1.5) and (1.3), obtained earlier by us. More precisely, we prove

Theorem 1. *If $p(z)$ is a polynomial of degree n, having no zeros in $|z| < 1$, then for β with $|\beta| \leq 1$*

$$\max_{|z|=1}\left|zp'(z) + \frac{n\beta}{2}p(z)\right| \leq \frac{n}{2}\left\{\left(\left|1 + \frac{\beta}{2}\right| + \left|\frac{\beta}{2}\right|\right)M(p, 1) - m\left(\left|1 + \frac{\beta}{2}\right| - \left|\frac{\beta}{2}\right|\right)\right\}, \tag{1.13}$$

$$\begin{aligned}\max_{|z|=1}\left|p(Rz) + \beta\left(\frac{R+1}{2}\right)^n p(z)\right| &\leq \frac{1}{2}\left\{\left|R^n + \beta\left(\frac{R+1}{2}\right)^n\right|\right.\\ &\quad \left. + \left|1 + \beta\left(\frac{R+1}{2}\right)^n\right|\right\}M(p, 1)\\ &\quad - \frac{m}{2}\left\{\left|R^n + \beta\left(\frac{R+1}{2}\right)^n\right| - \left|1 + \beta\left(\frac{R+1}{2}\right)^n\right|\right\}, \quad R > 1.\end{aligned} \tag{1.14}$$

Equality holds in (1.13) *and* (1.14) *for $p(z) = \lambda + \mu z^n$ with $|\lambda| \geq |\mu|$.*

Theorem 2. *If $p(z)$ is a polynomial of degree n, having all its zeros in $|z| \leq 1$, then for β with $|\beta| \leq 1$*

$$\max_{|z|=1}\left|zp'(z) + \frac{n\beta}{2}p(z)\right| \geq \frac{n}{2}[\{1 + \mathrm{Re}(\beta)\}M(p, 1) + m|\{1 + \mathrm{Re}(\beta)\} - |\beta||]. \tag{1.15}$$

Equality holds in (1.15) *for $p(z) = Ce^{i\alpha}z^n, C > 0$ and $\beta \geq 0$.*

Remark 1. Theorem 1 is a refinement of Theorem F, it can be easily seen by observing that

$$\left|1 + \frac{\beta}{2}\right| \geq \left|\frac{\beta}{2}\right|, \quad |\beta| \leq 1,$$

and

$$\left|R^n + \beta\left(\frac{R+1}{2}\right)^n\right| \geq \left|1 + \beta\left(\frac{R+1}{2}\right)^n\right|, \quad |\beta| \leq 1 \text{ and } R > 1.$$

Remark 2. Theorem 2 is a refinement of Theorem G.

2. Lemmas

For the proofs of the theorems, we require the following lemmas.

Lemma 1. *If $p(z)$ is a polynomial of degree n, having all its zeros in $|z| \leq 1$, then*

$$|p'(z)| \geq \frac{n}{2}|p(z)|, \qquad |z| = 1.$$

This lemma is due to Malik and Vong [6]. It suffices to observe that if $p(z) = c\Pi_{\nu=1}^{n}(z - z_\nu)$, then for $|z| = 1$, we have

$$R\left(\frac{zp'(z)}{p(z)}\right) = \sum_{\nu=1}^{n} R\left(\frac{z}{z - z_\nu}\right) \geq \frac{n}{2}.$$

Lemma 2. *If $p(z)$ is a polynomial of degree n, having all its zeros in $|z| \leq 1$, then*

$$|p(\mathrm{Re}^{i\theta})| \geq \left(\frac{R+1}{2}\right)^n |p(\mathrm{e}^{i\theta})|, \qquad R > 1 \quad \text{and} \quad 0 \leq 0 < 2\pi.$$

This lemma is due to Jain [4]. It was observed by Rivlin [10] that if f is a polynomial of degree at most n such that $f(z) \neq 0$ in $|z| < 1$, then

$$|f(\rho \mathrm{e}^{i\theta})| \geq \left(\frac{1+\rho}{2}\right)^n |f(\mathrm{e}^{i\theta})|, \quad (0 \leq \rho < 1, 0 \leq \theta < 2\pi).$$

Applying this result to the polynomial $f(z) := z^n\overline{p(1/\bar{z})}$ with $\rho := 1/R$ we obtain the desired estimate.

Lemma 3. *If $p(z)$ is a polynomial of degree n, having all its zeros in $|z| \leq 1$, then for β with $|\beta| \leq 1$*

$$\min_{|z|=1}\left|z\rho'(z) + \frac{n\beta}{z}p(z)\right| \geq mn\left|1 + \frac{\beta}{2}\right|, \tag{2.1}$$

$$\min_{|z|=1}\left|p(Rz) + \beta\left(\frac{R+1}{2}\right)^n\right| \geq m\left|R^n + \beta\left(\frac{R+1}{2}\right)^n\right|, \qquad R > 1. \tag{2.2}$$

Equality holds in (2.1) *and* (2.2) *for $p(z) = m\mathrm{e}^{i\gamma} z^n, m > 0$.*

Proof of Lemma 3. If $p(z)$ has a zero on $|z| = 1$, then inequalities (2.1) and (2.2) are trivial. Therefore we assume that $p(z)$ has all its zeros in $|z| < 1$. Then $m > 0$ and for α with $|\alpha| < 1$, we have

$$|\alpha m z^n| < m \leq |p(z)|, \qquad |z| = 1, \quad (\text{by}(1.9)),$$

thereby implying by Rouché's theorem that the polynomial

$$p_1(z) = p(z) - \alpha m z^n$$

has all its zeros in $|z| < 1$. On applying Lemma 1, we get

$$|z\{p'(z) - \alpha m n z^{n-1}\}| \geq \frac{n}{2}|p(z) - \alpha m z^n|, \quad |z| = 1 \quad \text{and} \quad |\alpha| < 1.$$

Therefore for $|\beta| < 1$ and $|\alpha| < 1$, the polynomial

$$z\{p'(z) - \alpha m n z^{n-1}\} + \beta\frac{n}{2}\{p(z) - \alpha m z^n\}$$

i.e.

$$\left\{zp'(z) + \frac{n\beta}{2}p(z)\right\} - \alpha nmz^n\left\{1 + \frac{\beta}{2}\right\}$$

will have no zeros on $|z| = 1$. As $|\alpha| < 1$, we have for β with $|\beta| < 1$

$$\left|zp'(z) + \frac{n\beta}{2}p(z)\right| \geq \left|nmz^n\left(1 + \frac{\beta}{2}\right)\right|, \quad |z| = 1,$$

i.e.

$$\left|zp'(z) + \frac{n\beta}{2}p(z)\right| \geq mn\left|1 + \frac{\beta}{2}\right|, \quad |z| = 1. \tag{2.3}$$

For β with $|\beta| = 1$, (2.3) follows by continuity. And now, the inequality (2.1) follows.

On applying Lemma 2 to the polynomial $p_1(z)$, we get for $R > 1$ and $|\alpha| < 1$

$$|p(Rz) - \alpha mR^n z^n| \geq \left(\frac{R+1}{2}\right)^n |p(z) - \alpha mz^n|, \quad |z| = 1.$$

Therefore for $|\beta| < 1$ and $|\alpha| < 1$, the polynomial

$$p(Rz) - \alpha mR^n z^n + \beta\left(\frac{R+1}{2}\right)^n \{p(z) - \alpha mz^n\},$$

i.e.

$$\left\{p(Rz) + \beta\left(\frac{R+1}{2}\right)^n p(z)\right\} - \alpha mz^n\left\{R^n + \beta\left(\frac{R+1}{2}\right)^n\right\}$$

will have no zeros on $|z| = 1$. As $|\alpha| < 1$, we have for β with $|\beta| < 1$

$$\left|p(Rz) + \beta\left(\frac{R+1}{2}\right)^n p(z)\right| \geq \left|mz^n\left\{R^n + \beta\left(\frac{R+1}{2}\right)^n\right\}\right|, \quad |z| = 1,$$

i.e.

$$\left|p(Rz) + \beta\left(\frac{R+1}{2}\right)^n p(z)\right| \geq m\left|R^n + \beta\left(\frac{R+1}{2}\right)^n\right|, \quad |z| = 1,$$

and the inequality (2.2) follows. This completes the proof of Lemma 3.

Lemma 4. *Let $Q(z)$ be a polynomial of degree n, having all its zeros in* $|z| \leq 1$ *and* $S(z)$ *be a polynomial of degree not exceeding that of* $Q(z)$. *If*

$$|S(z)| \leq |Q(z)| \tag{2.4}$$

for $|z| = 1$, *then for any* $|\beta| \leq 1$,

$$\left|\frac{zS'(z)}{n} + \beta\frac{S(z)}{2}\right| \leq \left|\frac{zQ'(z)}{n} + \beta\frac{Q(z)}{2}\right| \tag{2.5}$$

for $|z| = 1$.

This lemma is due to Malik and Vong [6]. However, this result is contained in ([9], Theorem 3.4) where it is shown that under the conditions of lemma 4,

$$|B_n S(z)| \leq |B_n Q(z)|, \quad (|z| = 1),$$

for every B_n-operator. It may be added that a linear operator T, which carries polynomials of degree at most n into polynomials of degree at most n, is called a B_n-operator provided that $T[f]$ has all its zeros in the open unit disc if f is of exact degree n and has all its zeros in the open unit disc.

Lemma 5. *If $p(z)$ is a polynomial of degree n, with $M(p,1) = 1$, then for $|\beta| \leq 1$ and $|z| = 1$*

$$\left|zp'(z) + \frac{n\beta}{2}p(z)\right| + \left|zq'(z) + \frac{n\beta}{2}q(z)\right| \leq n\left(\left|\frac{\beta}{2}\right| + \left|1 + \frac{\beta}{2}\right|\right),$$

where

$$q(z) = z^n\overline{p(1/\bar{z})}. \tag{2.6}$$

This lemma is due to Rahman ([8], inequality (5.3)).

Lemma 6. *Let $Q(z)$ be a polynomial of degree n, having all its zeros in $|z| \leq 1$. If $S(z)$ is a polynomial of degree at most n such that*

$$|S(z)| \leq |Q(z)|, \quad \textit{for} \quad |z| = 1, \tag{2.7}$$

then for β with $|\beta| \leq 1$ and $R \geq 1$, we have

$$\left|S(Rz) + \beta\left(\frac{R+1}{2}\right)^n S(z)\right| \leq \left|Q(Rz) + \beta\left(\frac{R+1}{2}\right)^n Q(z)\right|, \quad |z| = 1. \tag{2.8}$$

This lemma is due to Jain [4].

Lemma 7. *If $p(z)$ is a polynomial of degree at most n such that $M(p,1) = 1$, then for β with $|\beta| \leq 1, R \geq 1$ and $|z| = 1$*

$$\left|p(Rz) + \beta\left(\frac{R+1}{2}\right)^n p(z)\right| + \left|q(Rz) + \beta\left(\frac{R+1}{2}\right)^n q(z)\right|$$
$$\leq \left|1 + \beta\left(\frac{R+1}{2}\right)^n\right| + \left|R^n + \beta\left(\frac{R+1}{2}\right)^n\right|.$$

where $q(z)$ is, as in lemma 5.
This lemma is due to Jain [4].

Lemma 8. *If $p(z)$ is a polynomial of degree n, having all its zeros in $|z| \leq 1$, then for β with $|\beta| \leq 1$ and $|z| = 1$,*

$$\left|zp'(z) + \frac{n\beta}{2}p(z)\right| \geq \frac{n}{2}\{1 + \mathrm{Re}(\beta)\}|p(z)|.$$

This lemma is due to Jain ([3], Remark 2).

3. Proofs of the theorems

Proof of Theorem 1. If $p(z)$ has a zero on $|z| = 1$, then Theorem 1 reduces to ([3], Theorem 1). Therefore we assume that $p(z)$ has all its zeros in $|z| > 1$ (i.e. $m > 0$). Now for α with $|\alpha| < 1$, we have

$$|\alpha m| < m \leq |p(z)|, \quad |z| = 1, \quad \text{(by (1.9))},$$

thereby implying by Rouché's theorem that the polynomial

$$p_2(z) = p(z) - \alpha m$$

has no zeros in $|z| < 1$. Therefore the polynomial

$$\begin{aligned} q_2(z) &= z^n\overline{p_2(1/\bar{z})} \\ &= q(z) - \bar{\alpha} m z^n, \quad \text{(by (2.6))} \end{aligned}$$

will have all its zeros in $|z| \leq 1$. Also

$$|p_2(z)| = |q_2(z)|, \quad |z| = 1.$$

On applying Lemma 4, we get for $|z| = 1$,

$$\left|zp_2'(z) + \frac{n\beta}{2}p_2(z)\right| \leq \left|zq_2'(z) + \frac{n\beta}{2}q_2(z)\right|,$$

i.e.

$$\left|\{zp'(z) + \frac{n\beta}{2}p(z)\} - \frac{n\beta}{2}\alpha m\right| \leq \left|\left\{zq'(z) + \frac{n\beta}{2}q(z)\right\} - \bar{\alpha} mnz^n\left(1 + \frac{\beta}{2}\right)\right|, \quad |\alpha| < 1,$$

i.e.

$$\left|zp'(z) + \frac{n\beta}{2}p(z)\right| - nm|\alpha|\frac{|\beta|}{2} \leq \left|\left|zq'(z) + \frac{n\beta}{2}q(z)\right| - |\alpha|\, mn\left|1 + \frac{\beta}{2}\right|\right|, \quad |\alpha| < 1. \tag{3.1}$$

The polynomial $q(z)$, given by (2.6) has all its zeros in $|z| \leq 1$ and

$$\min_{|z|=1}|q(z)| = \min_{|z|=1}|p(z)| = m, \quad \text{(by (1.9))}.$$

And so, by Lemma 3 (inequality (2.1))

$$\min_{|z|=1}\left|zq'(z) + \frac{n\beta}{2}q(z)\right| \geq mn\left|1 + \frac{\beta}{2}\right|,$$

thereby allowing us to rewrite (3.1) as

$$\left|zp'(z) + \frac{n\beta}{2}p(z)\right| - mn|\alpha|\frac{|\beta|}{2} \leq \left|zq'(z) + \frac{n\beta}{2}q(z)\right| - |\alpha| mn\left|1 + \frac{\beta}{2}\right|, \quad |z| = 1 \quad \text{and} \quad |\alpha| < 1.$$

As $|\alpha| \rightarrow 1$, we get for $|z| = 1$,

$$\left|zp'(z) + \frac{n\beta}{2}p(z)\right| - \left|zq'(z) + \frac{n\beta}{2}q(z)\right| \leq -mn\left(\left|1 + \frac{\beta}{2}\right| - \left|\frac{\beta}{2}\right|\right). \tag{3.2}$$

Now, by lemma 5, we have for $|z| = 1$,

$$\left|zp'(z) + \frac{n\beta}{2}p(z)\right| + \left|zq'(z) + \frac{n\beta}{2}q(z)\right| \leq n\left(\left|1 + \frac{\beta}{2}\right| + \left|\frac{\beta}{2}\right|\right)M(p, 1). \tag{3.3}$$

Addition of inequalities (3.2) and (3.3) easily leads to inequality (1.13)

On applying lemma 6 to the polynomials $p_2(z)$ and $q_2(z)$, we get for $R > 1$ and $|z| = 1$

$$\left|p_2(Rz)+\beta\left(\frac{R+1}{2}\right)^n p_2(z)\right| \le \left|q_2(Rz)+\beta\left(\frac{R+1}{2}\right)^n q_2(z)\right|,$$

i.e.

$$\left|\left\{p(Rz)+\beta\left(\frac{R+1}{2}\right)^n p(z)\right\} - \alpha m\left\{1+\beta\left(\frac{R+1}{2}\right)^n\right\}\right|$$
$$\le \left|\left\{q(Rz)+\beta\left(\frac{R+1}{2}\right)^n q(z)\right\} - \bar{\alpha} m z^n\left\{R^n+\beta\left(\frac{R+1}{2}\right)^n\right\}\right|, \quad |\alpha|<1,$$

i.e.

$$\left|p(Rz)+\beta\left(\frac{R+1}{2}\right)^n p(z)\right| - m|\alpha|\left|1+\beta\left(\frac{R+1}{2}\right)^n\right|$$
$$\le \left|\left|q(Rz)+\beta\left(\frac{R+1}{2}\right)^n q(z)\right| - |\alpha|\, m\left|R^n+\beta\left(\frac{R+1}{2}\right)^n\right|\right|,$$
$$|\alpha| < 1. \qquad (3.4$$

Further on applying lemma 3 (inequality (2.2)) to the polynomial $q(z)$, we get for $R > 1$

$$\min_{|z|=1}\left|q(Rz)+\beta\left(\frac{R+1}{2}\right)^n q(z)\right| \ge \left|R^n+\beta\left(\frac{R+1}{2}\right)^n\right| \min_{|z|=1}|q(z)|,$$
$$= m\left|R^n+\beta\left(\frac{R+1}{2}\right)^n\right|,$$

thereby allowing us to rewrite (3.4) as

$$\left|p(Rz)+\beta\left(\frac{R+1}{2}\right)^n p(z)\right| - m|\alpha|\left|1+\beta\left(\frac{R+1}{2}\right)^n\right|$$
$$\le \left|q(Rz)+\beta\left(\frac{R+1}{2}\right)^n q(z)\right| - |\alpha| m\left|R^n+\beta\left(\frac{R+1}{2}\right)^n\right|,$$
$$|z| = 1, \quad R > 1 \text{ and } \quad |\alpha| < 1.$$

As $|\alpha| \to 1$, we get for $|z| = 1$ and $R > 1$,

$$\left|p(Rz)+\beta\left(\frac{R+1}{2}\right)^n p(z)\right| - \left|q(Rz)+\beta\left(\frac{R+1}{2}\right)^n q(z)\right|$$
$$\le -m\left\{\left|R^n+\beta\left(\frac{R+1}{2}\right)^n\right| - \left|1+\beta\left(\frac{R+1}{2}\right)^n\right|\right\}. \qquad (3.5$$

Now, by lemma 7, we have for $|z| = 1$ and $R > 1$,

$$\left|p(Rz)+\beta\left(\frac{R+1}{2}\right)^n p(z)\right| + \left|q(Rz)+\beta\left(\frac{R+1}{2}\right)^n q(z)\right|$$
$$\le \left\{\left|R^n+\beta\left(\frac{R+1}{2}\right)^n\right| + \left|1+\beta\left(\frac{R+1}{2}\right)^n\right|\right\} M(p,1). \qquad (3.6$$

Addition of inequalities (3.5) and (3.6) easily leads to inequality (1.14). This also completes the proof of Theorem 1.

Proof of Theorem 2. If $p(z)$ has a zero on $|z| = 1$, then Theorem 2 reduces to ([3], Remark 2). Therefore we assume that $p(z)$ has all its zeros in $|z| < 1$. Now as in the proof of lemma 3, for α with $|\alpha| < 1$, the polynomial

$$p_1(z) = p(z) - \alpha m$$

will have all its zeros in $|z| < 1$. On applying lemma 8, we get for α with $|\alpha| < 1$ and $|z| = 1$,

$$\left| zp'(z) + \frac{n\beta}{2}\{p(z) - \alpha m\} \right| \geq \frac{n}{2}\{1 + \text{Re}(\beta)\}|p(z) - \alpha m|, \tag{3.7}$$

i.e.

$$\left| \left| zp'(z) + \frac{n\beta}{2}p(z) \right| - \left| \frac{n\beta}{2}\alpha m \right| \right| \geq \frac{n}{2}\{1 + \text{Re}(\beta)\}\{|p(z)| - |\alpha m|\}. \tag{3.8}$$

Further, by lemma 3 (inequality (2.1)), we have for $|z| = 1$.

$$\begin{aligned} \left| zp'(z) + \frac{n\beta}{2}p(z) \right| &\geq nm\left|1 + \frac{\beta}{2}\right|, \\ &\geq nm\frac{|\beta|}{2}, \\ &\geq \left| \frac{n\beta}{2}\alpha m \right|, \quad \text{for} \quad |\alpha| < 1, \end{aligned}$$

thereby allowing us to rewrite (3.8) as

$$\left| zp'(z) + \frac{n\beta}{2}p(z) \right| - \left| \frac{n\beta}{2}\alpha m \right| \geq \frac{n}{2}\{1 + \text{Re}(\beta)\}\{|p(z)| - |\alpha m|\},$$

$$|z| = 1 \quad \text{and} \quad |\alpha| < 1.$$

As $|\alpha| \to 1$, we get for $|z| = 1$,

$$\left| zp'(z) + \frac{n\beta}{2}p(z) \right| \geq \frac{n}{2}\{1 + \text{Re}(\beta)\}|p(z)| + \frac{nm}{2}[|\beta| - \{1 + \text{Re}(\beta)\}],$$

thereby implying

$$\max_{|z|=1}\left| zp'(z) + \frac{n\beta}{2}p(z) \right| \geq \frac{n}{2}(\{1 + \text{Re}(\beta)\}M(p, 1) + m[|\beta| - \{1 + \text{Re}(\beta)\}]). \tag{3.9}$$

Again, by (3.7), we have for $|\alpha| < 1$ and $|z| = 1$,

$$\left| zp'(z) + \frac{n\beta}{2}p(z) \right| + \left| \frac{n\beta}{2}\alpha m \right| \geq \frac{n}{2}\{1 + \text{Re}(\beta)\}\{|p(z)| + |\alpha m|\}.$$

As $|\alpha| \to 1$, we get for $|z| = 1$,

$$\left| zp'(z) + \frac{n\beta}{2}p(z) \right| \geq \frac{n}{2}\{1 + \text{Re}(\beta)\}|p(z)| + \frac{nm}{2}[\{1 + \text{Re}(\beta)\} - |\beta|],$$

thereby implying

$$\max_{|z|=1}\left|zp'(z)+\frac{n\beta}{2}p(z)\right| \geq \frac{n}{2}(\{1+\mathrm{Re}(\beta)\}M(p,1)+m[\{1+\mathrm{Re}(\beta)\}-|\beta|]). \tag{3.10}$$

From (3.9) and (3.10), Theorem 2 follows.

References

[1] Ankeny N C and Rivlin T J, On a theorem of S. Bernstein, *Pacific J. Math.* **5** (1955) 849–852

[2] Aziz A and Dawood Q M, Inequalities for a polynomial and its derivative, *J. Approx. Theory* **54** (1988) 306–313

[3] Jain V K, Generalization of certain well known inequalities for polynomials, *Glas. Mat.* **32** (52) (1997) 45–51

[4] Jain V K, On maximum modulus of polynomials, *Ind. J. Pure Appl. Math.* **23(11)** (1992) 815–819

[5] Lax P D, Proof of a conjecture of P. Erdös on the derivative of a polynomial, *Bull Am. Math. Soc.* **50** (1944) 509–513

[6] Malik M A and Vong M C, Inequalities concerning the derivative of polynomials, *Rendiconti Del Circolo Matematico Di Palermo Serie II* **34** (1985) 422–426

[7] Polya G and Szegö G, Problems and theorems in Analysis, (Springer Verlag, Berlin, Heidelberg) (1972) Vol. 1

[8] Rahman Q I, Functions of exponential type, *Trans. Am. Math. Soc.* **135** (1969) 295–309

[9] Rahman Q I and Schmeisser G, Les inegalités de Markoff et de Bernstein, *Séminaire de Mathématiques Supérieures* (Les presses de l'Université de Montréal) (1983) No. 86

[10] Rivlin T J, On the maximum modulus of polynomials, *Am. Math. Mon.* **67** (1960) 251–253

[11] Schaeffer A C, Inequalities of A. Markoff and S. Bernstein for polynomials and related functions, *Bull. Am. Math. Soc.* **47** (1941) 565–579

[12] Turán P, Über die Ableitung von polynomen, *Compositio Math.* **7** (1939) 89–95

Proc. Indian Acad. Sci. (Math. Sci.), Vol. 110, No. 2, May 2000, pp. 147–155.
© Printed in India

Oscillations of first order difference equations

N PARHI

Department of Mathematics, Berhampur University, Berhampur 760 007, India

MS received 10 June 1999; revised 28 December 1999

Abstract. The oscillatory and asymptotic behaviour of solutions of first order difference equations is studied.

Keywords. Oscillation; disconjugacy; non-oscillation; difference equation; asymptotic behaviour.

1. Introduction

First order difference equations occur as mathematical models of some real world problems (see [1,4]). Although books on difference equations devote some space on such equations (see [3,4]), it seems that the qualitative behaviour of their solutions is not yet studied systematically. In ([4], (see p. 64)), a geometric approach is adopted to study asymptotic behaviour of solutions of a class of nonlinear homogeneous first order difference equations of the form

$$y_{n+1} = f(y_n).$$

The general solution of

$$z' + p(t)z = 0, \quad t \geq 0, \tag{1}$$

is given by

$$z(t) = A\exp\left(-\int_0^t p(s)\mathrm{d}s\right)$$

which has no zero in $[0,\infty)$ if $A \neq 0$. The discrete analogue of (1) is

$$y_{n+1} + p_n y_n = 0, \quad n \in [0,\infty) = \{0,1,2,\ldots\}, \tag{2}$$

which is oscillatory if $p_n > 0,\ n \geq 0$ (see Theorem 1 below). This is one of the properties which distinguishes difference equations from differential equations. In this paper we study oscillation of (2), the associated nonhomogeneous equation

$$y_{n+1} + p_n y_n = b_n \tag{3}$$

and the nonlinear equations

$$y_{n+1} + p_n G(y_n) = 0 \tag{4}$$

and

$$y_{n+1} + p_n G(y_n) = b_n. \tag{5}$$

By a solution of (2) on $[0,\infty)$ we mean a sequence $\{y_n\}$ of real numbers which satisfies (2) for $n \geq 0$. A solution $\{y_n\}$ of (2) is said to be nontrivial if for every $N \geq 0$ there exists $n \geq N$ such that $y_n \neq 0$. We may note that (3) (or (5)) does not admit a trivial solution if b_n is not the trivial sequence and (4) admits a trivial solution if $G(0) = 0$. By a solution of (2) (or (4)) we mean a nontrivial solution. Each of the equations (2)–(5) admits a unique solution if y_0 is given. A solution $\{y_n\}$ of (2) is said to be oscillatory if for every $N \geq 0$ there exists $n \geq N$ such that $y_{n-1}y_n \leq 0$; otherwise, $\{y_n\}$ is said to be non-oscillatory. Equation (2) is said to be oscillatory if every solution of (2) is oscillatory. It is said to be non-oscillatory if it admits a non-oscillatory solution. A sequence $\{y_n\}, n \geq 0$, of real numbers is said to have a generalized zero or node at $n = n_0$ provided that $y_{n_0} = 0$ if $n_0 = 0$ and if $n_0 > 0$, then $y_{n_0} = 0$ or $y_{n_0-1}y_{n_0} < 0$. Thus a solution $\{y_n\}$ of (2) is oscillatory if and only if the generalized zeros of $\{y_n\}$ are unbounded. Equation (2) is said to be disconjugate on $[0,\infty)$ if no solution of (2) has a generalized zero in $[0,\infty)$. The above definitions apply to eqs (3)–(5). The concept of generalized zeros or nodes was first introduced by Hartman in his work [2].

2. Oscillatory behaviour of solutions

In this section we study oscillatory/non-oscillatory behaviour of solutions of eqs (2)–(5). We assume

$$(\mathrm{H}_1) \qquad G \in C(R,R) \quad \text{with} \quad uG(u) > 0 \quad \text{for} \quad u \neq 0.$$

Remark. If $\{y_n\}$ is a solution of (2) with $y_0 = 0$, then it is a trivial solution. If $p_n = 0$ for some $n \geq 0$, then $y_m = 0$ for $m \geq n+1$ and hence $\{y_n\}$ is a trivial solution of (2). Thus we assume $y_0 \neq 0$ and $p_n \neq 0$ for $n \geq 0$ when we consider (2). From (H_1) it follows that $G(0) = 0$. Hence the above observation holds for eq. (4).

Theorem 1. (i) *If $p_n < 0$, $n \geq 0$, then eq.* (2) *is disconjugate on* $[0,\infty)$. (ii) *If $p_n > 0$, $n \geq 0$, then eq.* (2) *is oscillatory.* (iii) *If $\{p_n\}$ is oscillatory in the sense that for every $N \geq 0$ there exists $n \geq N$ such that $p_{n-1}\,p_n < 0$, then eq.* (2) *is oscillatory.*

Proof. Any solution of (2) is given by

$$y_n = (-1)^n y_0 \prod_{i=0}^{n-1} p_i. \tag{6}$$

(i) If $p_n < 0$ for $n \geq 0$, then writing (6) as

$$y_n = (-1)^{2n} y_0 \prod_{i=0}^{n-1} (-p_i),$$

we may observe that $y_n > 0$ or < 0 for $n \geq 0$ as $y_0 > 0$ or < 0. Thus (2) is disconjugate on $[0,\infty)$.

(ii) Since $p_n > 0$ for $n \geq 0$, we obtain from (6) that

$$y_{n-1}y_n = (-1)^{2n-1} y_0^2 p_{n-1} \prod_{i=0}^{n-2} p_i^2$$

for $n \geq 1$. Hence (2) is oscillatory.

(iii) Suppose that $\{p_n\}$ is oscillatory. Let $N \geq 0$. Choose $N^* \geq N+1$. Since $\{p_n\}$ is oscillatory, there exists $m \geq N^*$ such that $p_{m-1}\, p_m < 0$. If $p_m > 0$, for $m \geq N^*$, then from (6) we have

$$y_m y_{m+1} = (-1)^{2m+1} p_m\, y_0^2 \prod_{i=0}^{m-1} p_i^2 < 0.$$

If $p_m < 0$ for $m \geq N^*$, then $p_{m-1} > 0$ and hence

$$y_{m-1} y_m = (-1)^{2m-1} p_{m-1}\, y_0^2 \prod_{i=0}^{m-2} p_i^2 < 0.$$

Thus $\{y_n\}$ is oscillatory. Since it is an arbitrary solution of (2), eq. (2) is oscillatory.

Hence the theorem is proved.

Theorem 2. *Let* $(1+p_n) \geq 0$ *for* $n \geq 0$. *Suppose there exists a sequence* $\{c_n\}$ *with the following property:*

(H_2) *for every* $n \geq 0$ *there exists* $m \geq n$ *such that*

$$c_{m-1} c_m < 0 \quad \text{and} \quad b_n = c_{n+1} - c_n.$$

If

$$\sum_{n=0}^{\infty} (1+p_n) c_n^+ = +\infty \quad \text{and} \quad \sum_{n=0}^{\infty} (1+p_n) c_n^- = +\infty,$$

then eq. (3) *is oscillatory, where* $c_n^+ = \max\{c_n, 0\}$ *and* $c_n^- = \max\{-c_n, 0\}$.

Proof. Let $\{y_n\}, n \geq 0$, be a non-oscillatory solution of (3). Hence there exists $N \geq 0$ such that $y_n > 0$ or < 0 for $n \geq N$. Let $y_n > 0$ for $n \geq N$. Writing eq. (3) as

$$(y_{n+1} - c_{n+1}) - (y_n - c_n) + (1+p_n) y_n = 0, \tag{7}$$

we notice that the sequence $\{y_n - c_n\}$ is monotonic decreasing for $n \geq N$. We claim that $y_n - c_n > 0$ for $n \geq N$. If not, then $y_l - c_l \leq 0$ for some $l \geq N$. Then $k \geq l$ implies that $y_k - c_k \leq y_l - c_l \leq 0$ and hence $0 < y_k \leq c_k$ for $k \geq l$, a contradiction to (H_2). Hence our claim holds. From (7) we get

$$\begin{aligned}\sum_{n=N}^{N+i} (1+p_n) y_n &= \sum_{n=N}^{N+i} [(y_n - c_n) - (y_{n+1} - c_{n+1})] \\ &= (y_N - c_N) - (y_{N+i+1} - c_{N+i+1}) \\ &< y_N - c_N\end{aligned}$$

which implies that

$$\sum_{n=N}^{\infty} (1+p_n) y_n \leq y_N - c_N < +\infty.$$

On the other hand, $y_n \geq c_n^+$ for $n \geq N$ implies that

$$\sum_{n=N}^{\infty} (1+p_n) y_n \geq \sum_{n=N}^{\infty} (1+p_n) c_n^+ = +\infty,$$

a contradiction. Hence $y_n < 0$ for $n \geq N$. From (7) it follows that $\{y_n - c_n\}$ is monotonic increasing for $n \geq N$. If $y_l - c_l \geq 0$ for some $l \geq N$, then $k \geq l$ implies that $y_k - c_k \geq y_l - c_l \geq 0$, that is, $0 > y_k \geq c_k$ for $k \geq l$, which contradicts (H$_2$). Thus $y_n - c_n < 0$ for $n \geq N$. Since $-y_n \geq c_n^-$, then

$$\sum_{n=N}^{\infty}(1+p_n)y_n \leq -\sum_{n=N}^{\infty}(1+p_n)c_n^- = -\infty. \ ,$$

However, (7) yields

$$\sum_{n=N}^{\infty}(1+p_n)y_n \geq y_N - c_N > -\infty,$$

a contradiction. Hence $\{y_n\}$ is oscillatory. This completes the proof of the theorem.

Remark. We may note that $\{b_n\}$ changes sign, that is, for every $n \geq 0$ there exists $l \geq n$ such that $b_{l-1}b_l < 0$ if and only if (H$_2$) holds. Indeed, if $\{b_n\}$ changes sign, then defining $c_0 = 0, c_n = \sum_{i=0}^{n-1} b_i,\ n \geq 1$, we obtain $c_{n+1} - c_n = b_n$. Let $l \geq n+1$. If $b_l < 0$, then taking $m = l+1$, we get

$$c_{m-1}c_m = c_l c_{l+1} = \left(\sum_{i=0}^{l-1} b_i\right)^2 b_l < 0.$$

If $b_l > 0$, then $b_{l-1} < 0$. For $m = l$, we get

$$c_{m-1}c_m = c_{l-1}c_l = \left(\sum_{i=0}^{l-2} b_i\right)^2 b_{l-1} < 0.$$

Thus (H$_2$) is satisfied. Further, if (H$_2$) holds, then there exists $m \geq n+1$ such that $c_{m-1}c_m < 0$ and $c_m c_{m+1} < 0$. Hence $b_m = c_{m+1} - c_m$ implies that $b_{m-1}b_m < 0$, that is, $\{b_n\}$ is changing sign.

Theorem 3. *Let* $(1+p_n) \leq 0$ *for* $n \geq 0$. *Suppose there exists a sequence* $\{c_n\}$ *satisfying* (H$_2$). *If* $\sum_{n=0}^{\infty}(1+p_n)c_n^+ = -\infty$, $\sum_{n=0}^{\infty}(1+p_n)c_n^- = -\infty$ *and* $-\infty < \liminf_{n\to\infty} c_n < \limsup_{n\to\infty} c_n < \infty$, *then every solution of* (3) *oscillates or tends to* $\pm\infty$ *as* $n \to \infty$.

Proof. If possible, let $\{y_n\}$ be a non-oscillatory solution of (3). Hence $y_n > 0$ or < 0 for $n \geq N \geq 0$. Let $y_n > 0$ for $n \geq N$. From (7) it follows that the sequence $\{y_n - c_n\}$ is monotonic increasing for $n \geq N$. If possible, let there exist an $l \geq N$ such that $y_l - c_l = 0$. If $y_k - c_k = y_l - c_l$ for every $k \geq l$, then $c_k = y_k > 0$ for $k \geq l$, a contradiction to (H$_2$). Hence there exists a $k_1 > l$ such that $y_{k_1} - c_{k_1} > y_l - c_l = 0$. Then $n \geq k_1$ implies that $y_n - c_n \geq y_{k_1} - c_{k_1} > 0$. If such an l does not exist, then $y_n - c_n < 0$ or > 0 for $n \geq N$. In the former case, $0 < y_n < c_n$ for $n \geq N$, which contradicts (H$_2$). Hence $y_n - c_n > 0$ for $n \geq N$. Thus, in any case, there exists $k^* \geq N$ such that $y_n - c_n > 0$ for $n \geq k^*$. If $\lambda = \lim_{n\to\infty}(y_n - c_n)$, then $0 < \lambda \leq +\infty$. Suppose that $0 < \lambda < +\infty$. From (7) it follows that

$$-\sum_{n=k^*}^{\infty}(1+p_n)y_n \leq \lim_{i\to\infty}(y_{k^*+i} - c_{k^*+i}) = \lambda.$$

However, $y_n \geq c_n^+$ for $n \geq k^*$ implies that

$$-\sum_{n=k^*}^{\infty}(1+p_n)y_n \geq -\sum_{n=k^*}^{\infty}(1+p_n)c_n^+ = +\infty,$$

a contradiction. If $\lambda = +\infty$, then

$$\begin{aligned}\liminf_{n\to\infty} y_n &= \liminf_{n\to\infty}[(y_n - c_n) + c_n] \\ &\geq \liminf_{n\to\infty}(y_n - c_n) + \liminf_{n\to\infty} c_n \\ &= +\infty\end{aligned}$$

implies that $\lim_{n\to\infty} y_n = +\infty$. Similarly, if $y_n < 0$ for $n \geq N$, then we may show that $\lim_{n\to\infty} y_n = -\infty$. This completes the proof of the theorem.

The following examples illustrate the above results.

Example 1. Consider

$$y_{n+1} - \tfrac{1}{2}y_n = \tfrac{3}{2}(-1)^{n+1}, \quad n \geq 0.$$

Defining $c_0 = 1$ and

$$c_n = \begin{cases} 1, & n \text{ even} \\ -\frac{1}{2}, & n \text{ odd}, \end{cases}$$

we notice that $c_{n+1} - c_n = b_n$, $n \geq 0$ and $c_{n-1}c_n < 0$ for $n \geq 1$. As

$$c_n^+ = \begin{cases} 1, & n \text{ even} \\ 0, & n \text{ odd} \end{cases} \quad \text{and} \quad c_n^- = \begin{cases} 0, & n \text{ even} \\ \frac{1}{2}, & n \text{ odd} \end{cases}$$

then

$$\sum_{n=0}^{\infty}(1+p_n)c_n^+ = \sum_{i=0}^{\infty}(1+p_{2i})c_{2i}^+ = +\infty$$

and

$$\sum_{n=0}^{\infty}(1+p_n)c_n^- = \sum_{i=0}^{\infty}(1+p_{2i+1})c_{2i+1}^- = +\infty.$$

From Theorem 2 it follows that every solution of the equation oscillates. In particular, $\{y_n\} = \{(-1)^n\}$ is an oscillatory solution of the equation.

Example 2. Consider

$$y_{n+1} - 2y_n = 3(-1)^{n+1}, \quad n \geq 0.$$

Clearly, $b_n = c_{n+1} - c_n$, for $n \geq 0$, where

$$c_n = \begin{cases} 1, & n \text{ even} \\ -2, & n \text{ odd}, \end{cases}$$

$c_{n-1}c_n < 0$ for $n \geq 1$, and $-\infty < \liminf_{n\to\infty} c_n = -2 < \limsup_{n\to\infty} c_n = 1 < \infty$. Since

$$\sum_{n=0}^{\infty}(1+p_n)c_n^+ = \sum_{i=0}^{\infty}(1+p_{2i})c_{2i}^+ = -\infty$$

and

$$\sum_{n=0}^{\infty}(1+p_n)c_n^- = \sum_{i=0}^{\infty}(1+p_{2i+1})c_{2i+1}^- = -\infty,$$

then every solution of the equation oscillates or tends to ±∞ as $n \to \infty$ by Theorem 3. In particular, $\{y_n\} = \{(-1)^n\}$ is an oscillatory solution of the equation.

Remark. If $p_n \equiv -1$, then a solution of (3) is given by

$$y_n = y_0 + \sum_{i=0}^{n-1} b_i, \quad n \geq 1.$$

If $\{b_n\}$ changes sign, then it is not possible to conclude that eq. (3) is oscillatory. For example, we take $b_n = (-1)^n$, $n \geq 0$. Then $\{y_n\}$ is an oscillatory solution if $-1 \leq y_0 \leq 0$, a positive solution if $y_0 > 0$ and a negative solution if $y_0 < -1$. Thus the behaviour of solutions of $y_{n+1} - y_n = (-1)^n$, $n \geq 0$, is determind by the region of initial values of the solutions.

Remark. If $\{1 + p_n\}$ changes sign, then no information is available.

Remark. If $b_n \geq 0, p_n \leq 0$ with $p_n^2 + b_n^2 \neq 0, n \geq 0$, then eq. (3) admits a positive solution $\{y_n\}$ whenever $y_0 > 0$ and hence is non-oscillatory. However, if $b_n \geq 0, p_n \geq 0$ with $p_n^2 + b_n^2 \neq 0$, $n \geq 0$, then we cannot say that eq. (3) admits a non-oscillatory solution. Although several examples of the form (3) are given in [4] (see pp. 48–56), the qualitative behaviour of solutions is not studied.

The general solution of eq. (3) is given by (see p. 48, [4])

$$y_n = \left(\prod_{i=0}^{n-1}(-p_i)\right)\left[\sum_{r=0}^{n-1}\left(b_r \Big/ \prod_{i=0}^{r}(-p_i)\right) + A\right], \tag{8}$$

where $p_n \neq 0$ for $n \geq 0$ and A is an arbitrary constant. One may get different solutions of eq. (3) by changing A.

Theorem 4. *Let $p_n > 0$, $n \geq 0$. If*

$$\sum_{r=0}^{\infty}\left(b_r \Big/ \prod_{i=0}^{r}(-p_i)\right) = \pm\infty, \tag{9}$$

then every solution of eq. (3) *oscillates.*

Proof. It is possible to choose $N > 0$ sufficiently large such that

$$\sum_{r=0}^{n-1}\left(b_r \Big/ \prod_{i=0}^{r}(-p_i)\right) + A > 0 \text{ or } < 0$$

for $n \geq N$. Since

$$\prod_{i=0}^{n-1}(-p_i)\begin{cases} > 0, & \text{for } n \text{ odd} \\ < 0, & \text{for } n \text{ even}\end{cases}$$

then from (8) it follows that the generalized zeros of any solution $\{y_n\}$ of eq. (3) forms an unbounded set and hence is oscillatory. Thus the theorem is proved.

Example 3. Every solution of

$$y_{n+1} + y_n = b_n, \quad n \geq 0$$

oscillates, where

$$b_n = \begin{cases} 1, & n \text{ even} \\ 2^n, & n \text{ odd} \end{cases}$$

by Theorem 4, because

$$\sum_{r=0}^{\infty} \left(b_r \Big/ \prod_{i=0}^{r} (-p_i) \right)$$
$$= -1 + 2 - 1 + 2^3 - 1 + 2^5 - 1 + \cdots$$
$$\geq \sum_{r=0}^{\infty} 2^r = \infty.$$

Indeed, if $y_0 = A = 1$, then the generalized zeros of $\{y_n\}$ are at $n = 1, 2, 3, \ldots$.

Remark. If the series in (9) is oscillating, then eq. (3) may admit both oscillatory and non-oscillatory solutions.

Example 4. Consider

$$y_{n+1} + y_n = 1, n \geq 0.$$

Then the series

$$\sum_{r=0}^{\infty} \left(b_r \Big/ \prod_{i=0}^{r} (-p_i) \right) = \sum_{n=0}^{\infty} (-1)^{n+1}$$

is oscillating. Clearly, $y_0 = 1/2$,

$$y_n = (-1)^n \left[\sum_{r=0}^{n-1} (-1)^{r+1} + \frac{1}{2} \right], \quad n \geq 1,$$

is a positive solution of the equation and

$$u_n = (-1)^n \left[\sum_{r=0}^{n-1} (-1)^{r+1} + 2 \right], \quad n \geq 1 \quad \text{with} \quad u_0 = 2$$

is an oscillatory solution of the equation.

Remark. If the series in (9) is absolutely convergent, then eq. (3) may admit both oscillatory and non-oscillatory solutions.

Example 5. For the difference equation

$$y_{n+1} + y_n = \frac{1}{(n+1)^2}, \quad n \geq 0,$$

$$\sum_{r=0}^{\infty}\left|\left(b_r\Big/\prod_{i=0}^{r}(-p_i)\right)\right| = \sum_{r=0}^{\infty}\left|\frac{(-1)^{r+1}}{(r+1)^2}\right| = \sum_{r=1}^{\infty}\frac{1}{r^2} < \infty.$$

Using (8) we write

$$y_n = (-1)^n\left[\sum_{r=0}^{n-1}\frac{(-1)^{r+1}}{(r+1)^2} + A\right], n \geq 1.$$

If $A \geq 1$, then $\{y_n\}$ is an oscillatory solution of the equation with $y_0 = A$. If $A = 5/6$, then $\{y_n\}$ is a positive solution of the equation with $y_0 = 5/6$.

The asympotic behaviour of solutions of (4) is studied geometrically in ([4], p. 64). We have the following result concerning eq. (4). We may note that a solution of (4) cannot be expressed in the form (6).

Theorem 5. (i) *If* $p_n < 0$ *for* $n \geq 0$, *then* (4) *is disconjugate on* $[0, \infty)$. (ii) *If* $p_n > 0$ *for* $n \geq 0$, *then* (4) *is oscillatory.* (iii) *If* $\{p_n\}$ *is oscillatory, then* (4) *is oscillatory.*

Proof. (i) If possible, let $k \in [0, \infty) = \{0, 1, 2, \ldots\}$ be a generalized zero of a solution $\{y_n\}$ of (4). If $k = 0$, then $y_0 = 0$ and hence $\{y_n\}$ is a trivial solution. Let $k \in (0, \infty)$. If $y_k = 0$, then $y_n = 0$ for $n \geq k$, that is, $\{y_n\}$ is a trivial solution. If $y_k \neq 0$, then $y_{k-1}y_k < 0$. However, $y_k > 0$ implies that $y_{k-1} < 0$ and hence

$$0 < y_k = -p_{k-1}G(y_{k-1}) < 0,$$

a contradiction and $y_k < 0$ implies that $y_{k-1} > 0$ and hence $0 > y_k = -p_{k-1}G(y_{k-1}) > 0$, a contradiction. Thus $\{y_n\}$ has no generalized zero in $[0, \infty)$, that is, eq. (4) is disconjugate on $[0, \infty)$.

(ii) If possible, let $\{y_n\}$ be a non-oscillatory solution of (4). Hence $y_n > 0$ or < 0 for $n \geq N > 0$. Let $y_n > 0$ for $n \geq N$. From (4) we obtain

$$0 < y_{n+1} = -p_nG(y_n) < 0,$$

a contradiction. A similar contradiction is obtained if $y_n < 0$ for $n \geq N$. Then $\{y_n\}$ is oscillatory.

(iii) Let $\{y_n\}$ be a non-oscillatory solution of (4) such that $y_n > 0$ for $n \geq N > 0$. Let $N^* \geq N + 1$. Since $\{p_n\}$ changes sign, then there exists $k \geq N^*$ such that $p_{k-1}p_k < 0$. If $p_k > 0$, then we obtain

$$0 < y_{k+1} = -p_kG(y_k) < 0,$$

a contradiction. Suppose that $p_k < 0$. Then $p_{k-1} > 0$ and hence

$$0 < y_k = -p_{k-1}G(y_{k-1}) < 0,$$

a contradiction. We may obtain a similar contradiction if $y_n < 0$ for $n \geq N$. Thus eq. (4) is oscillatory.

This completes the proof of the theorem.

Theorem 6. (i) *If* $p_n \geq 0$ *and* $\{b_n\}$ *changes sign, then* (5) *is oscillatory.* (ii) *If* $p_n \leq 0$ *and* $b_n \geq 0$ *such that* $p_n^2 + b_n^2 > 0$, *then* (5) *is non-oscillatory.*

Proof. (i) If $\{y_n\}$ is a non-oscillatory solution of (5) with $y_n > 0$ for $n \geq N$, then (5) yields

$$0 < y_{n+1} + p_n G(y_n) = b_n,$$

a contradiction. A similar contradiction is obtained if $y_n < 0$ for $n \geq N$. Thus eq. (5) is oscillatory. (ii) If $y_0 > 0$, then $\{y_n\}$ is a positive solution of (5) and hence (5) is non-oscillatory.

Remark. The following examples suggest that if $\{b_n\}$ changes sign and either $p_n < 0$ or $\{p_n\}$ changes sign, then eq. (5) admits an oscillatory solution or is oscillatory.

Example 6. Clearly, $\{(-1)^n\}$ is an oscillatory solution of each of the following equations

$$y_{n+1} - 2y_n^3 = 3(-1)^{n+1}, \quad n \geq 0 \tag{10}$$

and

$$y_{n+1} + (-\tfrac{1}{2})^n y_n^3 = (-1)^{n+1}(1 + 2^{-n}), \quad n \geq 0.$$

We note that (10) admits a positive solution $\{y_n\}$ with $y_0 = 2$.

References

[1] Devaney R, *An Introduction to Chaotic Dynamical Systems* (California: Benjamin/Cummings) (1986)

[2] Hartman P, Difference equations: Disconjugacy, principal solutions, Green's functions, complete monotonicity, *Trans. Am. Math. Soc.* **246** (1978) 1–30

[3] Kelley W G and Peterson A C, *Difference equations: An introduction with applications* (New York: Academic Press Inc.) (1991)

[4] Mickens R E, *Difference equations* (New York: Van Nostrand Reinhold Co.) (1987)

Proc. Indian Acad. Sci. (Math. Sci.), Vol. 110, No. 2, May 2000, pp. 157–204.
© Printed in India

Continuous rearrangement and symmetry of solutions of elliptic problems

FRIEDEMANN BROCK

Department of Mathematics, Univeristy of Missouri-Columbia, Columbia MO 65211, USA

MS received 6 April 1999; revised 18 October 1999

Abstract. This work presents new results and applications for the continuous Steiner symmetrization. There are proved some functional inequalities, e.g. for Dirichlet-type integrals and convolutions and also continuity properties in Sobolev spaces $W^{1,p}$. Further it is shown that the local minimizers of some variational problems and the nonnegative solutions of some semilinear elliptic problems in symmetric domains satisfy a weak, 'local' kind of symmetry.

Keywords. Continuous symmetrization; integral inequality; Dirichlet-type integral; semilinear elliptic problem; symmetry of the solution.

1. Introduction

Consider a variational problem of the following form

$$\text{(P)} \qquad J(v) \equiv \int_\Omega (G(x, v, |\nabla v|) - F(x, v))\mathrm{d}x \longrightarrow \text{Stat.!}, \qquad v \in K, \tag{1.1}$$

where K is a closed subset of $W_0^{1,p}(\Omega)$, $p \geq 1$, and Ω is a domain in $\mathbb{R}^n$. The *nonnegative* minimizers of problems like (P) may describe stable ('ground') states of equilibria in plasma physics, heat conduction and chemical reactors (for examples see [Di, F, K1]). We ask for symmetries of the solutions of (P), if G, F and Ω have certain 'symmetries'.

A well-known result is the following. Let $v^\star$ denote the Schwarz symmetrization of v (i.e. the radially symmetric nonincreasing rearrangement). Assume that $\Omega = \mathbb{R}^n$, $G = G(|\nabla v|)$ and G is a nonnegative and convex function with $G(0) = 0$, $F = F(v)$ and F is continuous, and K contains only nonnegative functions and has the property that, if $v \in K$, then also $v^\star \in K$. Then

$$J(v^\star) \leq J(v). \tag{1.2}$$

If, in addition, problem (P) has a *unique global* minimizer u, then we can infer from (1.2) that $u = u^\star$. (Note that this means that u is radially symmetric nonincreasing, i.e.

$$u = u(|x|) \text{ and } u \text{ is nonincreasing in } r, \quad (r = |x|).)$$

However, if the global minimizer is not unique, then the question arises whether there could be equality in (1.2) if $v \neq v^\star$. Unfortunately this case cannot be excluded, as the following simple example shows (see [BZ]).

Example 1.1. For some $p \geq 1$, let

$$J(v) := \int_{\mathbb{R}^n} |\nabla v|^p \mathrm{d}x. \tag{1.3}$$

Then there are nonnegative smooth functions v with compact support which are *not* radially symmetric and satisfy $J(v) = J(v^\star)$. Their level sets $\{v > c\}$, $c > 0$, are nested, but nonconcentric balls, and the set $\{\nabla v = 0\}$ has nonempty interior, that is the graph of v has 'plateaus'.

Physically relevant are not only the global minima but also the *local* minima and *critical points* of (P). To show symmetry properties of these functions, the above argument fails, because in general the Schwarz symmetrization $v^\star$ is not close to v. Even though one expects symmetric solutions in many cases, there are again exceptions. Here is another typical example.

Example 1.2. *Semilinear problem for the p-Laplacian*: Let B be a ball in $\mathbb{R}^n$ with centre $0, f \in C(\mathbb{R}_0^+)$, $p > 1$, and let $u \in C^2(\overline{B})$ satisfy

$$\begin{aligned} &-\Delta_p u \equiv -\nabla(|\nabla u|^{p-2}\nabla u) = f(u), \quad u > 0 \quad \text{in } B,\\ &u = 0 \quad \text{on } \partial B. \end{aligned} \tag{1.4}$$

Note that the associated variational problem is

$$\int_B \left(\frac{1}{p}|\nabla v|^p - F(u)\right)\mathrm{d}x \longrightarrow \text{Stat.!}, \quad v \in W_0^{1,p}(B), \tag{1.5}$$

where

$$F(v) := \int_0^v f(z)\mathrm{d}z.$$

If $p = 2$ and f is smooth then it is well-known (see [GNN]) that

$$\begin{aligned} &u = u^\star \quad \text{and}\\ &(\partial u)/(\partial r) < 0 \quad \text{in } B \setminus \{0\}, \qquad (r = |x|). \end{aligned} \tag{1.6}$$

However, if $p > 2$ or if f is not smooth, then the conclusion (1.6) holds only under some additional assumptions. Below we give a short (but not complete) list of sufficient criteria for (1.6):

(i) $p = 2$ and $f = f_1 + f_2$, where f_1 is smooth and f_2 is increasing, [GNN];
(ii) $p = n$ and $f(v) > 0$ for $v > 0$, [KP], (see also [Lio1] for the case $p = n = 2$);
(iii) $f \in C^1(\mathbb{R}_0^+)$ and ∇u vanishes only at 0, [BaN];
(iv) $f \in C^1(\mathbb{R}_0^+)$ and $1 < p < 2$, [DamPa].

The proofs for (i), (iii) and (iv) use the so-called *moving plane method* which turned out to be a very powerful technique in proving symmetry results for positive solutions of semilinear elliptic problems in symmetric domains during the last two decades (see e.g. Se, GNN, BeN, Da, Dam, DamPa, SeZ). The moving plane technique makes essential use of the maximum principle for elliptic equations and exploits the invariance of the equation with respect to reflections. If the differential operator of the problem *degenerates* then the method is often applicable only under additional assumptions on the solution. This concerns, for instance, the p-Laplacian operator for $p > 2$ (compare the case (iii) above).

The result (ii) was proved by combining an isoperimetric inequality and a Pohozaev-type identity. However this method is not applicable if $p \neq n$.

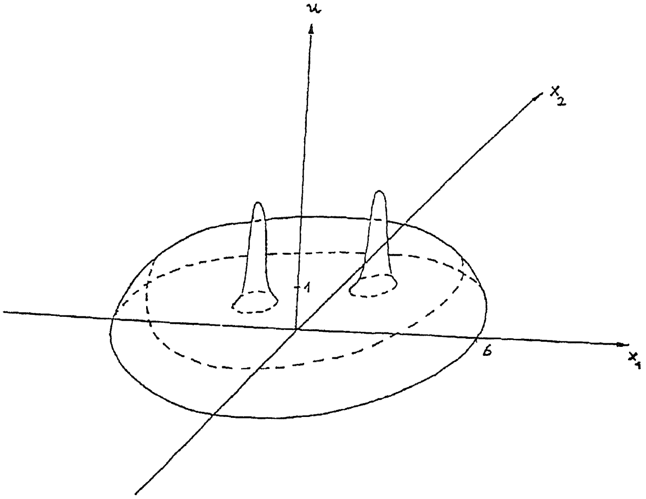


Figure 1.

One can construct radially symmetric solutions of (1.4) for which the second condition in (1.6) fails if either $p > 2$ and f is smooth, or if $p \in (1, 2]$ and f is Hölder continuous (see [GKPR]). Moreover, if $p = 2$ and f is only continuous and changes sign, then we cannot hope that the solution of (1.4) is radially symmetric. Below we give examples of solutions in the case $p \geq 2$ which have a plateau and two radially symmetric 'shifted bumps' on it. Note that similar examples can also be found in the recent paper [SeZ].

Let $p \geq 2, s > 2$,

$$w(x) = \begin{cases} (1 - |x|^2)^s & \text{if } |x| \leq 1 \\ 0 & \text{if } |x| > 1 \end{cases}, \quad \text{and}$$

$$v(x) = \begin{cases} 1 & \text{if } |x| < 5 \\ 1 - ((|x|^2 - 25)/11)^s & \text{if } 5 \leq |x| \leq 6 \end{cases}.$$

We choose $x^1, x^2 \in B_4$ with $|x^1 - x^2| \geq 2$ and set

$$u(x) := v(x) + w(x - x^1) + w(x - x^2) \qquad \forall x \in B_6.$$

The graph of u is built up by three radially symmetric 'mountains', one of them having a 'plateau' at height 1 while the other two are congruent to each other with their 'feet' lying on the plateau (see figure 1).

After a short computation we see that u is a solution of (1.4) with $\Omega = B_6$ and

$$f(u) := \begin{cases} (2s/11)^{p-1}(25 + 11(1-u)^{1/s})^{(p/2)-1}(1-u)^{p-(p/s)-1} \\ \quad \cdot \{(50/11)(p-1)(s-1) + (2ps - 2s - p + n)(1-u)^{1/s}\} \\ \hfill \text{if } 0 \leq u \leq 1, \\ (2s)^{p-1}(1 - (u-1)^{1/s})^{(p/2)-1}(u-1)^{p-(p/s)-1} \\ \quad \cdot \{-2(s-1)(p-1) + (2ps - 2s - p + n)(u-1)^{1/s}\} \\ \hfill \text{if } 1 \leq u \leq 2. \end{cases}$$

If $p=2$ and $s>2$ then we have $f\in C^{\infty}([0,2]\setminus\{1\})\cap C^{1-(2/s)}([0,2])$. The difference quotient of f is not bounded below near $u=1$, i.e. $f\notin C^1([0,2])$. In contrast, if $p>2$ and $s>p/(p-2)$, then we have $f\in C^1([0,2])$.

On the other hand, the functions in the above examples are distinguished by some 'local' symmetry which can be described as follows.

(LS) Every connected component of the subset

$$\{(x,u(x)) : 0<u(x)<\sup u,\ e\cdot\nabla u\neq 0\}$$

of the graph of u finds a congruent counterpart after reflection about some $(n-1)$-dimensional hyperplane $\{x : x\cdot e=\lambda\}$, $\lambda\in\mathbb{R}$, where e is some unit vector.

The purpose of this work is to obtain those weak symmetries for solutions of (P). The main analytic tool in the proofs will be some variant of continuous Steiner symmetrization which was developed in [B1]. Our approach is closely related to the corresponding variational problems of the differential equations. Therefore the applicability of the method seems to be restricted to equations in *divergence form*. On the other hand, we can also deal with *degenerate* elliptic operators. Furthermore, our regularity assumptions are rather mild. In most cases we only require that the solutions are differentiable in the interior of the domain and continuous up to the boundary, and the nonlinearity in the equation does not need to be smooth.

Given a Banach space X of measurable functions (e.g. $L^p(\mathbb{R}^n), p\in[1,+\infty)$), and a unit vector $e\in\mathbb{R}^n$, a *continuous Steiner symmetrization* is a continuous homotopy

$$t\longmapsto v^t,\quad 0\leq t\leq+\infty,$$

which connects $v\in X$ with its Steiner symmetrization in direction e, v^* (see Definition 2.6 and note the difference in notation to the Schwarz symmetrization $v^\star$), such that $v^0=v$ and $v^\infty=v^*$.

Clearly one looks for paths along which

$$J(v^t)\leq J(v),\qquad t\in[0,+\infty], \tag{1.7}$$

whenever

$$J(v^*)\leq J(v).$$

Bibliographical remarks on such homotopies were given in [B2]. Let us mention some related contributions which are connected with the *polarization* of a function or a set. This very simple kind of rearrangement was often used in the last decade to prove functional inequalities for symmetrizations (see e.g. [Du, Be, Ba] and the references cited therein).

Solynin [So] applied polarization methods to show that some capacities in the complex plane decrease under some type of continuous Steiner symmetrization. We mention that the same construction is much more simple for *convex* sets and was first used by McNabb [McN].

Finally, one can find a continuous *perturbation* of a given function (not just a homotopy!) which is formed by a certain scale of polarizations of this function (see [B3, BS]). This type of continuous rearrangement can be used to prove the symmetry of *local* minimizers for certain variational problems with potentials in a very simple manner.

Our variant of continuous symmetrization is a semigroup and satisfies the family of inequalities (1.7) for a large number of functionals, in particular for the Dirichlet-type integrals (1.3). In addition, it allows the following characterization of locally symmetric functions (see Theorem 6.2 for a more general formulation).

Let B be a ball in $\mathbb{R}^n$ with centre 0 and $p \in (1, +\infty)$. Further let $u \in W_0^{1,p}(B) \cap C^1(\overline{B})$ and $u \geq 0$. Then, if

$$\int_B (|\nabla u^t|^p - |\nabla u|^p)\mathrm{d}x = o(t) \qquad \text{as } t \searrow 0, \tag{1.8}$$

u satisfies the symmetry property (LS).

The symmetry proofs in this work depend on a number of technical steps. Let us explain the main line by considering the model equation (1.4).

Step 1. By multiplying (1.4) with $(u^t - u)$ and then by integrating we obtain

$$\int_B |\nabla u|^{p-2} \nabla u \nabla (u^t - u)\mathrm{d}x = \int_B f(u)(u^t - u)\mathrm{d}x. \tag{1.9}$$

(Note that, if $u \in W_0^{1,p}(B)$ is nonnegative, then the symmetrized functions u^t, $t \in [0, +\infty]$, also belong to $W_0^{1,p}(B)$ (see § 3), so that $(u^t - u)$ is an admissible function.)

Step 2. One shows that the right-hand side of (1.9) is of order $o(t)$ as $t \searrow 0$ (see §§ 4 and 5).

Step 3. By convexity the left-hand side of (1.9) is less than or equal to

$$\frac{1}{p}\int_B (|\nabla u^t|^p - |\nabla u|^p)\mathrm{d}x,$$

and this integral is less than or equal to 0 by (1.7) (see § 3). This yields (1.8), and so u is locally symmetric (see § 6).

Now we give an outline of our paper. In § 2 we give a new definition of the variant of continuous symmetrization which was investigated in [B2]. This new definition appears to be more transparent than the old one since it already contains the main properties of the continuous rearrangement, namely equimeasurability, monotonicity and the semi-group property. We show that open (compact) sets are transformed into open (respectively compact) sets under continuous symmetrization. At the end we recall some of the inequalities and continuity properties that we have derived in [B2] and which we will frequently use in our proofs.

The following §§ 3, 4 and 5 deal with properties of symmetrized functions in Sobolev spaces $W^{1,p}$. Most of these results are needed for the proofs of our symmetry theorems but a few of them are of independent interest in the theory of rearrangements. Those readers who are mostly interested in applications might skip these sections and return to them later. In § 3 we prove inequalities which compare some weighted norm of a nonnegative function $u \in W^{1,p}(\mathbb{R}^n)$ with the same norm of u^t. Note that similar inequalities are known for Steiner symmetrization and for the so-called starshaped rearrangements (see [K1, K4, BM] and [B4]). The proof is based on an approximation argument with a special dense subclass of piecewise smooth functions (called 'good' functions). These functions have the property that they oscillate only finitely often along any straight line lying in the direction of the symmetrization. In § 4 we show that continuous Steiner symmetrization is continuous from the right with respect to the parameter t in Sobolev spaces $W^{1,p}(\mathbb{R}^n)$, $1 \le p < +\infty$. In § 5 we study the behaviour of some nonlinear integral functionals for $t \searrow 0$ and show that it is approximately linear. In § 6 we investigate locally symmetric functions (see property (LS) above). A purely *analytic* description in terms of continuous Steiner symmetrization is given by Theorem 6.2. The preceding investigations enable us to prove that local minimizers – and also the corresponding weak solutions – of problem (P) are locally symmetric in 'symmetric' situations (Theorems 7.1–7.3 of § 7).

We point out that it is possible in many cases to derive from the *local* symmetry *additional* symmetries such as Steiner symmetry or radial symmetry (see [B4, B6]). Some further results in this direction will be published in a forthcoming paper.

2. Preliminaries

We introduce some notation. Let $\mathbb{R}^n$ be the Euclidean space, $\mathbb{R}_0^+ = [0, +\infty)$ and $\mathbb{R}^+ = (0, +\infty)$. If $n \ge 2$ and $x \in \mathbb{R}^n$, then we write

$$x = (x', y), \quad x' = (x_1, \dots, x_{n-1}), \quad y = x_n,$$

and $|x|$ for the norm of x. $B_r(x_0)$ denotes the open ball in $\mathbb{R}^n$ with radius r centered at x_0, and we write $B_r = B_r(0)$. By ω_n we denote the volume of the n-dimensional unit ball in $\mathbb{R}^n$. For any set M in $\mathbb{R}^n$ we denote with $\overline{M}$ its closure and with $\chi(M)$ its characteristic function. If A, B are two open or compact sets then $A + B := \{z : z = x + y, x \in A, y \in B\}$ denotes their Minkowski sum. Let $\mathcal{M}(\mathbb{R}^n)$ be the set of Lebesgue measurable – measurable in short – sets in $\mathbb{R}^n$ with *finite* measure. If $M \in \mathcal{M}(\mathbb{R}^n)$ then we denote by $|M|$ its n-dimensional measure and by $S(M) = (S_1(M), \dots, S_n(M))$ the centre of gravity where

$$S_i(M) = |M|^{-1} \int_M x_i \, \mathrm{d}x, \quad i = 1, \dots, n.$$

We write $M \Delta N$ for the symmetric difference $(M \setminus N) \cup (N \setminus M)$ of two measurable sets M and N. Generally we treat measurable sets only in *a.e. sense*, i.e. we write

$$\begin{aligned} &M = N \Longleftrightarrow |M \Delta N| = 0 \quad \text{and} \\ &M \subset N \Longleftrightarrow |M \setminus N| = 0. \end{aligned}$$

If Ω is an open set in $\mathbb{R}^n$ and $p \in [1, +\infty]$ then we denote by $\|\cdot\|_p$ the usual norm in the space $L^p(\Omega)$. Sometimes we will write

$$\|u\|_{p,G} := \begin{cases} \left(\displaystyle\int_G |u|^p \mathrm{d}x\right)^{1/p} & \text{if } 1 \le p < +\infty \\ \operatorname*{ess\,sup}_G |u| & \text{if } p = +\infty \end{cases},$$

to indicate the integration over a *subset* G of Ω. By $W^{1,p}(\Omega)$ we denote the Sobolev space of functions $u \in L^p(\Omega)$ having generalized partial derivatives $u_{x_i} \in L^p(\Omega)$, $i = 1, \dots, n$, and we write

$$\|u\|_{W^{1,p}(\Omega)} := \|u\|_p + \sum_{i=1}^n \|u_{x_i}\|_p$$

for the norm in this space. By $W_0^{1,p}(\Omega)$ we denote the completion of $C_0^\infty(\Omega)$ under the norm $\|\cdot\|_{W^{1,p}(\Omega)}$. Recall that $W_0^{1,p}(\mathbb{R}^n) = W^{1,p}(\mathbb{R}^n)$ (see [A]). By $C_0^{0,1}(\Omega)$ we denote the space of Lipschitzean functions with compact support in Ω. For any function space the subscript '+' denotes the corresponding subspace of nonnegative functions, e.g. $L^p_+(\Omega)$, $W^{1,p}_{0+}(\Omega)$, $C^{0,1}_{0+}(\Omega), \dots$.

A function $F : \mathbb{R}^+_0 \to \mathbb{R}^+_0$ is called a Young function if F is continuous and convex and if $F(0) = 0$. Finally, let $\mathcal{S}(\mathbb{R}^n)$ denote the class of real measurable functions u satisfying

$$|\{x \in \mathbb{R}^n : u(x) > c\}| < +\infty \quad \forall c > \inf u.$$

Note that $L^p_+(\mathbb{R}^n)$ and $W^{1,p}_+(\mathbb{R}^n)$, $1 \le p < +\infty$, are subspaces of $\mathcal{S}_+(\mathbb{R}^n)$.

Next we give the definitions of some well-known symmetrizations.

(1) Let $M \in \mathcal{M}(\mathbb{R})$, and let M be *open or compact*. Then set

$$M^* := \begin{cases} (-(1/2)|M|, +(1/2)|M|) & \text{if } M \text{ is open} \\ [-(1/2)|M|, +(1/2)|M|] & \text{if } M \text{ is compact and } M \neq \emptyset. \end{cases} \tag{2.1}$$

If $M \in \mathcal{M}(\mathbb{R})$ is *neither open nor compact*, then M^* is given by the first formula in (2.1) in a.e. sense. M^* is called the symmetrization of M.

If $u \in \mathcal{S}(\mathbb{R})$ then the function

$$u^*(x) := \begin{cases} \sup\{c > \inf u : x \in \{u > c\}^*\} & \text{if } x \in \bigcup\limits_{c > \inf u} \{u > c\}^* \\ \inf u & \text{if } x \notin \bigcup\limits_{c > \inf u} \{u > c\}^* \end{cases} \tag{2.2}$$

is called the *symmetrization* or the *symmetric nonincreasing rearrangement* of u.

Note that $u(x)$ is symmetric with respect to zero, nonincreasing for $x > 0$, and we have

$$\{u > c\}^* = \{u^* > c\} \quad \forall c > \inf u. \tag{2.3}$$

(2) Let $n \ge 2$ and $M \in \mathcal{M}(\mathbb{R}^n)$. For every $x' \in \mathbb{R}^{n-1}$ we set

$$M(x') := \{y \in \mathbb{R} : (x', y) \in M\}, \quad (\text{intersection of } M \text{ with } (x', \mathbb{R})).$$

Note that every set $M \in \mathcal{M}(\mathbb{R}^n)$ has the representation

$$M = \{x = (x', y) : y \in M(x'),\ x' \in \mathbb{R}^{n-1}\},$$

where $M(x') \in \mathcal{M}(\mathbb{R})$ for almost every $x' \in \mathbb{R}^{n-1}$. The set

$$M^* := \{x = (x', y) :\ y \in (M(x'))^*, x' \in \mathbb{R}^{n-1}\} \tag{2.4}$$

is called the *Steiner symmetrization* of M with respect to y. Note that M^* is symmetric and convex with respect to the hyperplane $\{y = 0\}$. Moreover the sets $(M(x'))^*$ and thus also M^* are *pointwise* given by formula (2.4) if M is open or compact. Also it is well-known that if M is open (respectively compact) then M^* is again open (respectively compact).

If $u \in \mathcal{S}(\mathbb{R}^n)$, then the function

$$u^*(x', y) := \begin{cases} \sup\{c > \inf u : y \in \{u(x', \cdot) > c\}^*\} & \text{if } y \in \bigcup\limits_{c > \inf u} \{u(x', \cdot) > c\}^* \\ \inf u & \text{if } y \notin \bigcup\limits_{c > \inf u} \{u(x', \cdot) > c\}^* \end{cases} \tag{2.5}$$

is called the *Steiner symmetrization* of u. Note that $u^*(x',y)$ is symmetric with respect to $\{y=0\}$, nonincreasing in y for $y>0$, and we have

$$\{u(x',\cdot)>c\}^* = \{u^*(x',\cdot)>c\} \text{ for } c>\inf u \text{ and } x' \in \mathbb{R}^{n-1}. \tag{2.6}$$

(3) Let M as in (2) and let $r>0$ satisfy $|M| = |B_r| = \omega_n r^n$. If M is open or compact, then set

$$M^\star := \begin{cases} B_r & \text{if } M \text{ is open} \\ \overline{B_r} & \text{if } M \text{ is compact and } M \neq \emptyset. \end{cases} \tag{2.7}$$

(Notice the difference between $M^\star$ and M^*!)

If M is *neither open nor compact* then $M^\star$ is given by the first formula in (2.7) in the a.e. sense. $M^\star$ is called the *Schwarz symmetrization* of M.

If $u \in \mathcal{S}(\mathbb{R}^n)$ then the function

$$u^\star(x) := \begin{cases} \sup\{c>\inf u : x \in \{u>c\}^\star\} & \text{if } x \in \bigcup\limits_{c>\inf u} \{u>c\}^\star \\ \inf u & \text{if } x \notin \bigcup\limits_{c>\inf u} \{u>c\}^\star \end{cases} \tag{2.8}$$

is called the *Schwarz symmetrization* or the *(radially) symmetric decreasing rearrangement* of u. Note that $u^\star$ can be written as $u^\star = u^\star(|x|)$ and is nonincreasing in $|x|$, and we have

$$\{u>c\}^\star = \{u^\star > c\} \quad \forall c > \inf u. \tag{2.9}$$

Further let us mention that for *continuous* functions u the level sets in (2.3), (2.6) and (2.9) are open such that the corresponding symmetrizations of u are *pointwise* given by these formulas. Also it is well-known that these symmetrizations are then continuous, too. Clearly for *measurable* functions the identities (2.3), (2.6) and (2.9) still hold in a.e. sense (and (2.6) for a.e. $x' \in \mathbb{R}^{n-1}$).

Remark 2.1. It is more convenient in the literature to define the symmetrizations of arbitrary measurable sets and functions *pointwise* (see e.g. [K1]). (For instance in case of the Steiner symmetrization this can be achieved by agreeing that $u^*(x',y)$ is right- (or left-) continuous in y for $y>0$.) But it will turn out that we cannot give a *pointwise* definition of *continuous* symmetrization for *arbitrary* measurable sets and functions. Since the Steiner symmetrization will appear in that context as a special case we prefered the above settings.

This paper deals with a variant of continuous Steiner symmetrization which was introduced by the author in [B2]. Below we give a new and much shorter definition:

DEFINITION 2.1

Continuous symmetrization of sets in $\mathcal{M}(\mathbb{R})$: A family of set transformations

$$E_t : \mathcal{M}(\mathbb{R}) \longrightarrow \mathcal{M}(\mathbb{R}), \qquad 0 \le t \le +\infty,$$

satisfying the properties ($M, N \in \mathcal{M}(\mathbb{R})$, $0 \le s,t \le +\infty$)

(i) $|E_t(M)| = |M|$, (equimeasurability),
(ii) If $M \subset N$, then $E_t(M) \subset E_t(N)$, (monotonicity),

(iii) $E_t(E_s(M)) = E_{s+t}(M)$, (semigroup property),

(iv) If $I = [y_1, y_2]$ is a bounded closed interval, then $E_t(I) = [y_1^t, y_2^t]$, where

$$\begin{aligned} y_1^t &= \tfrac{1}{2}(y_1 - y_2 + \mathrm{e}^{-t}(y_1 + y_2)), \\ y_2^t &= \tfrac{1}{2}(y_2 - y_1 + \mathrm{e}^{-t}(y_1 + y_2)), \end{aligned} \tag{2.10}$$

is called a continuous symmetrization.

Remark 2.2. One immediately verifies that the rules for the formation of symmetrized intervals (2.10) are consistent with (i)–(iii). Note also that there are possible other variants of the continuous symmetrization by modification of the formulas (2.10) (see [K2]). Some of the results in the following sections 2–4 could be proved similarly using these modified definitions. The present variant of continuous symmetrization can be used to give another *analytic* description of the symmetry property (LS) (see Theorem 6.2) which plays a central role in our approach. We underline that Theorem 6.2 is not true for the continuous symmetrization of [K2] in view of the examples given in ([B2], Remark 9). Therefore we will concentrate ourselves upon the present version.

From now on we will write for simplicity $M^t := E_t(M)$ for the symmetrized sets.

Theorem 2.1. *There exists a family of set transformations E^t, $0 \le t \le +\infty$, satisfying* (i)–(iv). *For every $M \in \mathcal{M}(\mathbb{R})$ the map $t \longmapsto M^t$, $0 \le t \le +\infty$, is a homotopy, i.e.*

$$M^0 = M, \quad M^\infty = M^*. \tag{2.11}$$

Finally, if $M \in \mathcal{M}(\mathbb{R})$ is open, then M^t has an open representative for every $t \in [0, +\infty]$.

Proof. First note that the properties (i) and (ii) imply

$$(M \cup N)^t \supset M^t \cup N^t, \tag{2.12}$$

$$(M \cap N)^t \subset M^t \cap N^t \quad \text{and} \tag{2.13}$$

$$|M \Delta N| \ge |M^t \Delta N^t|. \tag{2.14}$$

Now the proof is in several steps. Our aim is to give an explicit construction of the sets M^t, $t \in [0, +\infty]$, and to show the uniqueness of this construction.

(1) Let M be simple, that is $M = \cup_{k=1}^m I_k$, where the I_k's are disjoint bounded closed intervals. From (2.12) it follows that we must have

$$M^t \supset \bigcup_{k=1}^m I_k^t \qquad \forall t \in [0, +\infty].$$

The intervals I_k^t are disjoint for

$$t \le t_1 := \min\left\{ \log \frac{2|S(I_j) - S(I_k)|}{|I_j| + |I_k|} : 1 \le j,\ k \le m \right\},$$

and for $t = t_1$ some of them meet each other in their endpoints. In view of the equimeasurability (i) we must therefore have

$$M^t = \bigcup_{k=1}^m I_k^t \quad \text{for } 0 \le t \le t_1. \tag{2.15}$$

Furthermore, since the family M^t, $0 \leq t \leq +\infty$, must satisfy the semigroup property (iii), we can argue analogously for parameters $t \geq t_1$ by using the formula $M^t = (M^{t_1})^{t-t_1}$. Thus we get by induction numbers $m =: m_0 > m_1 > \cdots > m_{N-1} := 1$ and $0 =: t_0 < t_1 < \cdots < t_N := +\infty$, and bounded closed intervals $I_{k,l}$, $k = 1, \ldots, m_l$, such that for any $t \in [t_l, t_{l+1}]$ and any $l \in \{0, \ldots, N-1\}$

$$M^t = \bigcup_{k=1}^{m_l} (I_{k,l})^{t-t_l},$$

where the intervals $(I_{k,l})^{t-t_l}$ are pairwise disjoint for $t < t_{l+1}$, and where some of them coalesce for $t = t_{l+1}$. Moreover

$$|M^{t_{l+1}} \Delta M^t| \longrightarrow 0 \quad \text{as} \quad t \nearrow t_{l+1}, \qquad l = 0, \ldots, N-1.$$

Finally (2.11) is satisfied. Vice versa, it is easy to see that the above construction yields a family of set transformations which satisfies (i)–(iv) in the subclass of simple sets. Furthermore, by using the rule (2.14) we check that this construction is unique. Note also, that since we may add arbitrary nullsets to the sets M^t – the above representations remain unchanged if the I_k's are *open* bounded intervals.

(2) Let M be open and $t \in [0, +\infty]$. Then we have $M = \cup_{k=1}^{+\infty} I_k$, where the I_k's are open, pairwise disjoint intervals. Setting $M_m := \cup_{k=1}^{m} I_k$, $m = 1, 2, \ldots$, we must then have $M^t \supset \cup_{m=1}^{+\infty} M_m^t$ by (2.12). (Note that the sets M_m^t are well-defined by part (1)!) Since $|M_m| = |M_m^t| \longrightarrow |M|$ as $m \to +\infty$, and since (ii) must be fulfilled, this leads to

$$M^t = \bigcup_{m=1}^{+\infty} M_m^t. \tag{2.16}$$

By using (2.14) and part (1), we check easily, that the family M^t, $0 \leq t \leq +\infty$, given by (2.16) does not depend on the enumeration of the intervals I_k.

Vice versa, by using again (2.14), we see that the above construction satisfies all the properties (i)–(iv) in the subclass of open sets, and that this construction is unique. In particular, formula (2.16) shows that M^t has an open representative and that (2.11) is again satisfied.

(3) Let $M \in \mathcal{M}(\mathbb{R})$. Then we have a representation

$$M = \bigcap_{n=1}^{+\infty} O_n, \tag{2.17}$$

where $O_n \supset O_{n+1}$, $n = 1, 2, \ldots$, are open sets. To satisfy (2.13), we must also have $M^t \subset \cap_{n=1}^{+\infty} O_n^t$. On the other hand, we have that $O_n^t \supset O_{n+1}^t$, $n = 1, 2, \ldots$, that is $|O_n^t| \longrightarrow |M|$ as $n \to +\infty$. The rule (i) forces the following representation,

$$M^t = \bigcap_{n=1}^{+\infty} O_n^t. \tag{2.18}$$

In view of (2.14) and part (2) we see that the set M^t given by (2.18) is independent of the representation (2.14). Furthermore, by using part (2) and once more formula (2.14), we check that the above construction satisfies the rules (i)–(iv) in the class $\mathcal{M}(\mathbb{R})$, and that this construction is the only one, satisfying these properties. Finally, (2.11) is satisfied by part (2) and (2.18). ■

Theorem 2.1 enables us to give a *pointwise* definition of the continuous symmetrization of *open* sets.

DEFINITION 2.2

Continuous symmetrization of open sets in $\mathcal{M}(\mathbb{R})$: Let $M \in \mathcal{M}(\mathbb{R})$ be open and $t \in [0, +\infty]$. Then the set

$$M^{t,O} := \bigcup \{U : U \text{ is an open representative of } N^t,\ N \text{ open},\ N \subset\subset M\} \tag{2.19}$$

is called the precise (open) representative of M^t.

One verifies easily that the above definitions of continuous symmetrization on the real axis are equivalent to those given in [B2]. Next we repeat the definition of the continuous Steiner symmetrization of [B2].

DEFINITION 2.3

Continuous (Steiner) symmetrization of sets in $\mathcal{M}(\mathbb{R}^n)$: Let $M \in \mathcal{M}(\mathbb{R}^n)$, $n \geq 2$. Then the family of sets

$$M^t := \{x = (x', y) : y \in (M(x'))^t,\ x' \in \mathbb{R}^{n-1}\}, \quad 0 \leq t \leq +\infty, \tag{2.20}$$

is called the continuous Steiner symmetrization of M. If M is open and $t \in [0, +\infty]$, then the set

$$M^{t,O} := \{x = (x', y) : y \in (M(x'))^{t,O},\ x' \in \mathbb{R}^{n-1}\} \tag{2.21}$$

is called the precise representative of M^t. Here the relation "=" in (2.21) has to be understood in the pointwise sense.

Remark 2.3. (1) Note that if $M \in \mathcal{M}(\mathbb{R}^n)$, then we have by the above definition $M^{0,O} = M$ and $M^{\infty,O} = M^*$ in the pointwise sense. (2) According to ([B2], Theorem 4) the properties listed in Theorem 2.1 remain valid for continuous Steiner symmetrization ($n \geq 2$). Below we give three further properties:

(a) If $M, N \in \mathcal{M}(\mathbb{R}^n)$ are open sets with $M \subset N$, then

$$\operatorname{dist}\{M; \partial N\} \leq \operatorname{dist}\{M^{t,O}; \partial N^{t,O}\} \qquad \forall 0 \leq t \leq +\infty. \tag{2.22}$$

It is easy to verify (compare also [BS]) that (2.22) yields the following:
(b) *Smoothing property:* If $M \in \mathcal{M}(\mathbb{R}^n)$, $t \in [0, +\infty]$ and $r > 0$, then

$$M^{t,O} + B_r \subset (M + B_r)^{t,O} \qquad \text{and} \tag{2.23}$$

$$M^{t,O} \setminus (\partial M^{t,O} + \overline{B_r}) \supset (M \setminus (\partial M + \overline{B_r}))^{t,O}. \tag{2.24}$$

From the Definitions 2.1–2.3 we immediately derive the following property:
(c) *Continuity from the inside:* If $\{M_k\}$ is an increasing sequence of open sets with $|\bigcup_{k=1}^{+\infty} M_k| < +\infty$, then

$$\bigcup_{k=1}^{+\infty} (M_k)^{t,O} = \left(\bigcup_{k=1}^{+\infty} M_k\right)^{t,O} \qquad \forall t \in [0, +\infty]. \tag{2.25}$$

Theorem 2.1 suggests that if $M \in \mathcal{M}(\mathbb{R}^n)$ is open then the precise representatives $M^{t,($ $t \in [0, +\infty]$, should be open too. It was kindly pointed out to me by Buttazzo that this fa is missing in [B2]. Nevertheless its proof is simple and requires no more than th monotonicity (Definition 2.1(ii)) and the properties (b) and (c) from Remark 2.3. Th observation was first made by Sarvas [Sa] in a context of general rearrangements.

Lemma 2.1. *Let $M \in \mathcal{M}(\mathbb{R}^n)$ and open. Then the sets $M^{t,O}$, $0 \leq t \leq +\infty$, are open to*

Proof. We fix $t \in [0, +\infty]$. In view of (2.25) we have

$$M^{t,O} = \bigcup_{k=1}^{+\infty} (M \setminus (\partial M + (1/k)\overline{B_1}))^{t,O}.$$

By the monotonicity (Definition 2.1(ii)) and by (2.24) this yields

$$M^{t,O} = \bigcup_{k=1}^{+\infty} (M^{t,O} \setminus (\partial M^{t,O} + (1/k)\overline{B_1})),$$

which means that $M^{t,O}$ is open. The lemma is proved. ∎

Remark 2.4. Similarily as in the case of the Steiner and Schwarz symmetrization it is als possible to give the continuous symmetrization of *compact sets* a pointwise meaning (se [B4]). But since we do not need such a construction in this paper we omit the details.

From now on let us agree that if we speak about the continuous symmetrization of *ope* sets, then we always mean their precise representatives, and we omit the superscript *(*

DEFINITION 2.4

Continuous (Steiner) symmetrization of functions: Let $u \in \mathcal{S}(\mathbb{R}^n)$. Then the family functions u^t, $0 \leq t \leq +\infty$, defined by

$$u^t(x) := \begin{cases} \sup\{c > \inf u : x \in \{u > c\}^t\} & \text{if } x \in \bigcup\limits_{c > \inf u} \{u > c\}^t \\ \inf u & \text{if } x \notin \bigcup\limits_{c > \inf u} \{u > c\}^t \end{cases}, \quad x \in \mathbb{R}^n, \tag{2.2}$$

is called continuous (Steiner) symmetrization of u with respect to y in the case $n \geq 2$ ar continuous symmetrization in the case $n = 1$.

Remark 2.5. It is easy to see that formula (2.26) is equivalent to the following relatio

$$\begin{aligned} \{u^t > c\} &= \{u > c\}^t \qquad \forall c > \inf u, \\ \{u^t = \inf u\} &= \mathbb{R}^n \setminus \bigcup_{c > \inf u} \{u > c\}^t, \\ \{u^t = +\infty\} &= \bigcap_{c > \inf u} \{u > c\}^t. \end{aligned} \tag{2.2}$$

It was shown in [B2], that $u^0 = u$ and $u^\infty = u^*$. Furthermore, if u is continuous, then f every $t \in [0, +\infty]$ the function u^t has a continuous representative which is given by tl formulas (2.26) and (2.27) in pointwise sense and on the right-hand sides of the

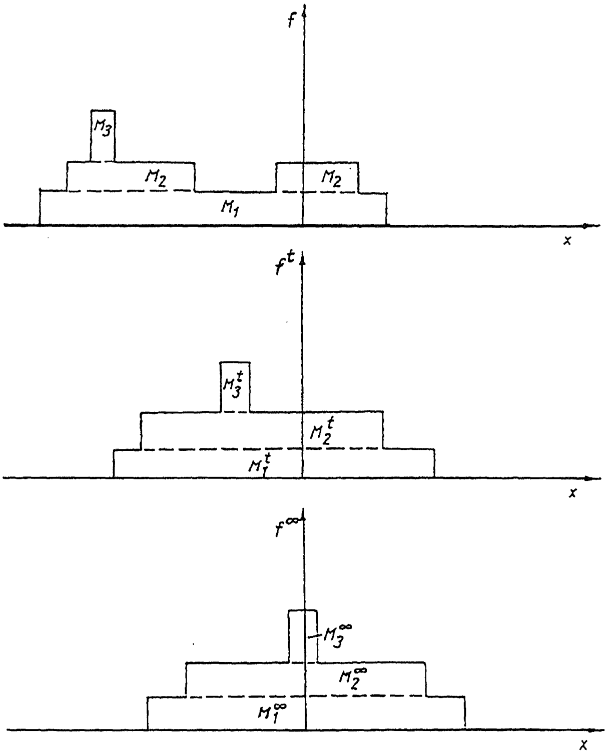


Figure 2.

formulas are taken the precise (open) representatives of the corresponding level sets. One can illustrate the formulas (2.26), (2.27) by continuously rearranging a step function as in figure 2.

If

$$u = c_0 + \sum_{i=}^{m} c_i \chi(M_i), \tag{2.28}$$

where $M_1 \supset \cdots \supset M_m$, $M_i \in \mathcal{M}(\mathbb{R}^n)$, and $c_0 \in \mathbb{R}$, $c_i > 0$, $i = 1, \ldots, m$, then

$$u^t = c_0 + \sum_{i=1}^{m} c_i \chi(M_i^t), \qquad t \in [0, +\infty]. \tag{2.29}$$

From now on let us agree that if we speak about the continuous symmetrizations of *continuous* functions, then we always mean their precise (continuous) representatives.

Remark 2.6. Let us recall some properties of continuous symmetrization that we proved in [B2] and which we will use from time to time $(M, N \in \mathcal{M}(\mathbb{R}^n), u, v, w \in \mathcal{S}_+(\mathbb{R}^n),$ $t \in [0, +\infty])$.

(1) *Monotonicity* (see [B2], Theorem 5):

If $u \leq v$, then

$$u^t \leq v^t. \tag{2.30}$$

(2) *Cavalieri's principle* (see [B2], Theorem 8):

$$\int_{\mathbb{R}^n} F(u)\mathrm{d}x = \int_{\mathbb{R}^n} F(u^t)\mathrm{d}x, \qquad (2.3$$

if F is Borel measurable and the left-hand side of (2.31) converges.

(3) *Continuity with respect to the parameter t:* If $t_m \to t$ as $m \to +\infty$, then (see [B2 Theorem 3)

$$M^{t_m} \longrightarrow M^t \qquad \text{in measure,} \qquad (2.3$$

and if u is a.e. finite, then (see [B2], Theorem 7)

$$u^{t_m} \longrightarrow u^t \qquad \text{in measure.} \qquad (2.3$$

(4) *Centre-formula* (see [B2], Remark 4):

$$S(M^t) = (S_1(M), \dots, S_{n-1}(M), e^{-t}S_n(M)). \qquad (2.3$$

(5) *Nonexpansivity in $L^p(\mathbb{R}^n)$*, $1 \le p \le +\infty$, (see [B2], Lemma 3): If $u, v \in L^p(\mathbb{R}^n)$, th

$$\|u^t - v^t\|_p \le \|u - v\|_p. \qquad (2.3$$

(6) *Hardy–Littlewood inequality* (see [B2], Lemma 4): If $u, v \in L^2(\mathbb{R}^n)$, then

$$\int_{\mathbb{R}^n} u^t v^t \,\mathrm{d}x \ge \int_{\mathbb{R}^n} uv \,\mathrm{d}x. \qquad (2.3$$

(7) If u is Lipschitz continuous with Lipschitz constant L then u^t is Lipschitz continuou too, with Lipschitz constant less or equal to L, (see [B2], Theorem 7).

Note that (1), (2), (5) and (6) are common properties of monotone equimeasurab rearrangements (see [K1, BS]). We mention that the Lipschitz continuity is in fact t 'best' regularity which is preserved under continuous symmetrization. This can be se by symmetrizing a function $f \in C^1(\mathbb{R})$ which has more than two monotonicity interv (see figure 3). The functions f^t and f^∞ are *not differentiable* in the marked points.

From now on we will assume that $n \ge 2$. Since the continuous Steiner symmetrizati is in fact a 'one-dimensional' construction, the results of this work can be transferred the simpler case $n = 1$ with obvious changes.

3. Dirichlet-type inequalities

In this section we prove various inequalities which compare some (weighted) Sobo norm of a function u with the same norm of u^t. The strategy in the proofs consists changing locally the variable of integration from y to u in the functionals. Functions which this is possible are characterized by the following:

DEFINITION 3.1 ('Good' functions)

A function u is called good if u is defined on $\mathbb{R}^n$ and nonnegative, piecewise smooth w compact support, if for every $x' \in \mathbb{R}^{n-1}$ and $c > 0$ the equation $u(x', y) = c$ has onl finite number of solutions $y = y_k$, $k = 1, \dots, l$, and if

$$\inf\{|u_y(x)| : x \in \mathbb{R}^n, u_y(x) \text{ exists}\} > 0. \qquad (3$$

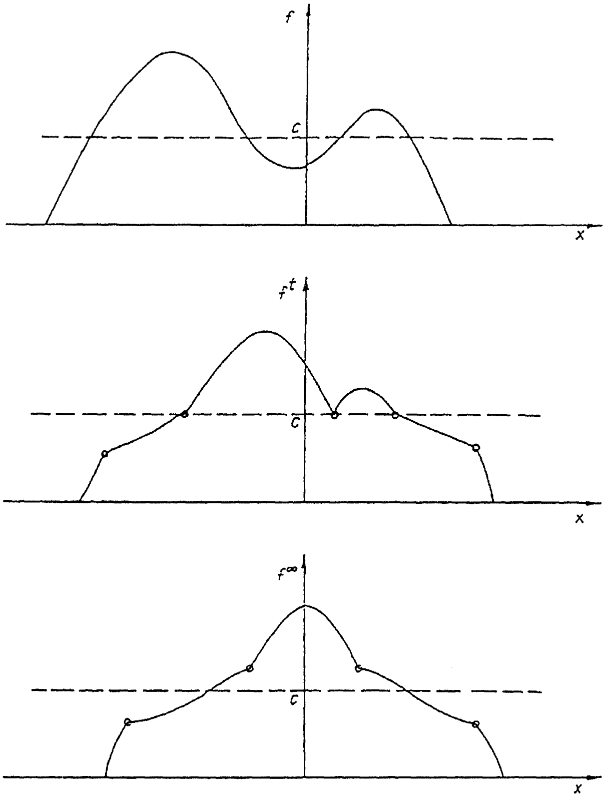


Figure 3.

Remark 3.1. (1) Good functions are dense in $W_+^{1,p}(\mathbb{R}^n)$ in the norm of $W^{1,p}(\mathbb{R}^n)$ for every $p \in [1, +\infty)$. This can be seen as follows. $C_0^\infty(\mathbb{R}^n)$ is dense in $W^{1,p}(\mathbb{R}^n)$. Any C_0^∞-function can be approximated by piecewise linear functions with compact support. If u is piecewise linear with support in $B_R(0)$, $(R > 0)$, then set

$$\varepsilon_0 := \min\{|u_y(x)| : x \in \mathbb{R}^n, u_y(x) \text{ exists and is } \neq 0\} > 0$$

and

$$v(x) := \begin{cases} 1 - R^{-1}|y| & \text{if } |y| < R \\ 0 & \text{if } |y| \geq R \end{cases}.$$

The functions $u_\varepsilon := u + \varepsilon v$ are good if $0 < |\varepsilon| < \varepsilon_0$, and u_ε converges to u in $W^{1,p}(\mathbb{R}^n)$ as ε tends to zero. Note that the same argumentation can be found in ([K1], pp. 49) for good piecewise linear functions in $W_0^{1,p}(\Omega)$, where Ω is a bounded domain. (2) Let $u \in W_+^{1,1}(\mathbb{R}^n) \cap C_{\text{loc}}^1(\mathbb{R}^n)$. Then u is absolutely continuous on almost every line $\{x' = \text{const}\}$ (see [EG], p. 164). From this we can infer (compare [C], Appendix 1 and 4) that u is 'generically' good, i.e. for *almost every* pair $(x', c) \in \mathbb{R}^{n-1} \times \mathbb{R}_0^+$ the equation $u(x', y) = c$ has only a finite number of solutions, and the equation $u_y(x', y) = 0$ does not have any solution. (3) If u is good and piecewise linear, then the functions u^t, $t \in [0, +\infty]$, are in general not piecewise linear, as one can see from simple examples of functions which are not quasiconcave in the direction y. But there holds the following:

Lemma 3.1. *Let u be good. Then the functions u^t, $t \in [0,+\infty]$, are good, too.*

Proof. The set

$$K := \{(x',u) \in \mathbb{R}^{n-1} \times \mathbb{R}^+ : \exists (x',y) \in \operatorname{supp} u, \text{ such that } u = u(x',y)\} \tag{3.2}$$

is compact. Furthermore, it is easy to see that for a.e. point $(x'_0, u_0) \in K$ there exists an open neighbourhood $V \subset K$ such that the equation $u = u(x',y)$ has *exactly* $2m$, $(m = m(V))$, solutions $y = y_k(x',u)$ in V, $y_k \in C^1(V)$, $k = 1, \ldots, 2m$, and such that $y_1 < \cdots < y_{2m}$. Thus u can be represented in V by *local* inverse functions $y = y_k(x',u)$, and we have that

$$u_y(x',y_k) = \left(\frac{\partial y_k}{\partial u}\right)^{-1} \begin{cases} > 0 & \text{if } k \text{ is odd} \\ < 0 & \text{if } k \text{ is even} \end{cases},$$
$$u_{x_i}(x',y_k) = -\frac{\partial y_k}{\partial x_i}\left(\frac{\partial y_k}{\partial u}\right)^{-1}, \quad i = 1, \ldots, n-1, \quad (k = 1, \ldots, 2m). \tag{3.3}$$

Our aim is to derive analogous representations for u^t, $t \in [0,+\infty]$. To this end we restrict our considerations to one of the above open sets V. First observe that for each $(x',u) \in V$ the equation $u = u^t(x',y)$ has at most $2m$ solutions y, by the proof of Theorem 2.1. Let V' be an open set with $V' \subset\subset V$. Then we see from Definitions 2.1–2.3 that for small t, u^t can be represented in V' by smooth inverse functions $y = y^t_k(x',u)$ through the formulas

$$y^t_{2k-1} = \frac{1}{2}(y_{2k-1} - y_{2k} + \mathrm{e}^{-t}(y_{2k-1} + y_{2k})),$$
$$y^t_{2k} = \frac{1}{2}(y_{2k} - y_{2k-1} + \mathrm{e}^{-t}(y_{2k-1} + y_{2k})), \tag{3.4}$$

and there hold the following identities,

$$u^t_y(x',y^t_k) = \left(\frac{\partial y^t_k}{\partial u}\right)^{-1} \begin{cases} > 0 & \text{if } k \text{ is odd} \\ < 0 & \text{if } k \text{ is even} \end{cases},$$
$$u^t_{x_i}(x',y^t_k) = -\frac{\partial y^t_k}{\partial x_i}\left(\frac{\partial y^t_k}{\partial u}\right)^{-1}, \quad i = 1, \ldots, n-1, \quad (k = 1, \ldots, 2m). \tag{3.5}$$

Suppose that $t_1 (= t_1(x',u))$ is the first value of t, such that some of the intervals $[y^t_{2k-1}(x',u), y^t_{2k}(x',u)]$, $k = 1, \ldots, m$, coalesce. Note that $t_1(x',u)$ varies continuously in V'. For simplicity in notation let us assume that we have $y^{t_1}_{2k} = y^{t_1}_{2k+1}$, $k = 1, \ldots, l$, $(l \leq m-1)$, at some point $(x'_0, u_0) \in V'$. Following the proof of Theorem 2.1, we see that there is some (small) neighbourhood V'' of (x'_0, u_0) and some number $t_2 > \sup\{t_1(x',u): (x',u) \in V''\}$, such that the functions $y^t_1(x',u)$ and $y^t_{2l}(x',u)$, $(t \leq t_1(x',u))$, find 'continuations' $\eta^t_i(x',u)$, for $t_1(x',u) < t < t_2$, and such that $u = u^t(x',\eta^t_i)$, $i = 1, 2$, in V''. From the equimeasurability we have that

$$\eta^t_2 - \eta^t_1 = \sum_{k=1}^{l}(y_{2k-1} - y_{2k}) = y^{t_1}_{2l} - y^{t_1}_1. \tag{3.6}$$

Since the set $\cup_{k=1}^{l}[y_{2k-1}, y_{2k}]$ is continuously symmetrized independently from the other intervals for $t < t_2$, we may apply the centre formula (2.34) onto this set. Together with the semigroup property this yields

$$\frac{\eta_1^t + \eta_2^t}{2} = \mathrm{e}^{t-t_1} \frac{\sum_{k=1}^{l} \frac{1}{2}(y_{2k}^{t_1} + y_{2k-1}^{t_1})}{\sum_{k=1}^{l}(y_{2k}^{t_1} - y_{2k-1}^{t_1})}. \tag{3.7}$$

After a differentiation with respect to x_i, $i = 1, \ldots, n-1$, we infer from (3.6), (3.7) that

$$\begin{aligned} \eta_{2,x_i}^t - \eta_{1,x_i}^t &= \sum_{k=1}^{l}(y_{2k,x_i}^{t_1} - y_{2k-1,x_i}^{t_1}), \\ \eta_{2,x_i}^t + \eta_{1,x_i}^t &= \frac{2\sum_{k=1}^{l}(y_{2k}^{t_1} y_{2k,x_i}^{t_1} - y_{2k-1}^{t_1} y_{2k-1,x_i}^{t_1})}{\sum_{k=1}^{l}(y_{2k}^{t_1} - y_{2k-1}^{t_1})} \\ &\quad - \frac{\sum_{k=1}^{l}((y_{2k}^{t_1})^2 - (y_{2k-1}^{t_1})^2)}{(\sum_{k=1}^{l}(y_{2k}^{t_1} - y_{2k-1}^{t_1}))^2} \sum_{k=1}^{l}(y_{2k,x_i}^{t_1} - y_{2k-1,x_i}^{t_1}). \end{aligned} \tag{3.8}$$

In view of the equalities

$$\begin{aligned} &y_{2k-1}^{t_1} = y_{2k}^{t_1}, \quad k = 1, \ldots, l, \\ &y_{2l}^{t_1} = \eta_2^{t_1}, \quad y_1^{t_1} = \eta_1^{t_1}, \end{aligned} \tag{3.9}$$

we obtain from (3.8)

$$\begin{aligned} \eta_{1,x_i}^{t_1} &= -\frac{1}{y_{2l}^{t_1} - y_1^{t_1}} \sum_{j=1}^{2l}(-1)^j y_{j,x_i}^{t_1}(y_{2l}^{t_1} - y_j^{t_1}), \\ \eta_{2,x_i}^{t_1} &= \frac{1}{y_{2l}^{t_1} - y_1^{t_1}} \sum_{j=1}^{2l}(-1)^j y_{j,x_i}^{t_1}(y_j^{t_1} - y_1^{t_1}). \end{aligned} \tag{3.10}$$

Analogously we compute the derivatives with respect to u as

$$\begin{aligned} |\eta_{1,u}^{t_1}| &= \frac{1}{y_{2l}^{t_1} - y_1^{t_1}} \sum_{j=1}^{2l} |y_{j,u}^{t_1}|(y_{2l}^{t_1} - y_1^{t_1}), \\ |\eta_{2,u}^{t_1}| &= \frac{1}{y_{2l}^{t_1} - y_1^{t_1}} \sum_{j=1}^{2l} |y_{j,u}^{t_1}|(y_j^{t_1} - y_1^{t_1}), \end{aligned} \tag{3.11}$$

and we have $\eta_{2,u}^{t_1} < 0 < \eta_{1,u}^{t_1}$. We will not specify formulas for the 'future' of the remaining intervals $[y_{2k-1}^{t_1}(x', u), y_{2k}^{t_1}(x', u)]$, $k = l+1, \ldots, m$, for $t \in (t_1(x', u), t_2)$. Some of these intervals might be computed henceforth according to (3.4) while others coalesce during that time. This leads to analogous computations.

By means of the semigroup property we can repeat these considerations step by step for every $t \in [0, +\infty]$ and for almost every points of K. From the formulas (3.4), (3.5), (3.10) and (3.11) we see that u^t is piecewise smooth and

$$\inf\{|u_y^t(x)| : x \in \mathbb{R}^n,\ u_y^t(x) \text{ exists}\} > 0, \quad t \in (0, +\infty].$$

The lemma is proved. ■

Theorem 3.1. *Let be u a good function, G a Young function and $a \in C(\mathbb{R})$ nonnegativ even and convex. Then for every $t \in [0, +\infty]$*

$$\int_{\mathbb{R}^n} G\left(\left\{a^2(y)\left(\frac{\partial u}{\partial y}\right)^2 + \sum_{i=1}^{n-1}\left(\frac{\partial u}{\partial x_i}\right)^2\right\}^{1/2}\right) \mathrm{d}x$$
$$\geq \int_{\mathbb{R}^n} G\left(\left\{a^2(y)\left(\frac{\partial u^t}{\partial y}\right)^2 + \sum_{i=1}^{n-1}\left(\frac{\partial u^t}{\partial x_i}\right)^2\right\}^{1/2}\right) \mathrm{d}x. \qquad \text{(3.1}$$

Proof. We use the notations of the previous proof. We may change locally the variable integration in (3.12) from (x', y) to (x', u). Then the integrals on the left and right-ha side of (3.12) become $\int_K I(x', u)\mathrm{d}x'\mathrm{d}u$ and $\int_K I_t(x', u)\mathrm{d}x'\mathrm{d}u$, respectively, where K is giv by (3.2) and I and I_t are nonnegative functions which will be specified below. Then, prove (3.12), it is sufficient to show that

$$I(x', u) \geq I_t(x', u) \quad \text{for a.e. } (x', u) \in K. \qquad \text{(3.1}$$

Using the notations of the previous proof, we compute

$$I(x_0', u_0) = \sum_{k=1}^{2m} G\left(\left(\left|\frac{\partial y_k}{\partial u}\right|\right)^{-1}\left\{a^2(y_k) + \sum_{i=1}^{n-1}\left(\frac{\partial y_k}{\partial x_i}\right)\right\}^{1/2}\right)\left|\frac{\partial y_k}{\partial u}\right|. \qquad \text{(3.1}$$

Similarly, we have that

$$I_t(x_0', u_0) = \sum_{k=1}^{2m} G\left(\left(\left|\frac{\partial y_k^t}{\partial u}\right|\right)^{-1}\left\{a^2(y_k^t) + \sum_{i=1}^{n-1}\left(\frac{\partial y_k^t}{\partial x_i}\right)\right\}^{1/2}\right)\left|\frac{\partial y_k^t}{\partial u}\right|, \quad \text{for } t \in (0, t_1$$
$$\text{(3.1}$$

where $t_1 = t_1(x_0', u_0)$. Thus, to prove (3.13) at (x_0', u_0) for $t \in (0, t_1)$, it suffices to show th

$$\varphi_k(t) := \sum_{l=2k-1}^{2k} G\left(\left(\left|\frac{\partial y_l^t}{\partial u}\right|\right)^{-1}\left\{a^2(y_l^t) + \sum_{i=1}^{n-1}\left(\frac{\partial y_l^t}{\partial x_i}\right)^2\right\}^{1/2}\right)\left|\frac{\partial y_l^t}{\partial u}\right|$$
$$\text{is nondecreasing for } t \in [0, t_1), \ (k = 1, \dots m). \qquad \text{(3.1}$$

To see this, we formally extend the definition (3.4) of the functions y_k^t, $(k = 1, \dots, 2n$ for all $t \in [0, +\infty]$. We introduce the new parameter $\lambda := (1/2)(1 - \mathrm{e}^{-t})$, and s $\psi_k(\lambda) := \varphi_k(t)$. By setting in addition

$$\psi_k(1 - \lambda) := \psi_k(\lambda) \qquad \forall \lambda \in [0, (1/2)],$$

a simple calculation shows that $\psi_k(\lambda)$, $\lambda \in [0, 1]$, is convex. This proves (3.16).

Next assume that at the moment $t = t_1$ the intervals $[y_{2k-1}^t, y_{2k}^t]$, $k = 1, \dots, l$, $(l \leq n$ coalesce and are 'continued' in a single interval $[\eta_1^t, \eta_2^t]$ according to the formulas (3. (3.7). Note that $I_t(x_0', u_0)$ is not defined at $t = t_1$. Setting $y_k := y_k^{t_1}$, $k = 1, \dots, 2l$, a $\eta_k := \eta_k^{t_1}$, $k = 1, 2$, we want to show that

$$\sum_{k=1}^{2} G\left((|\eta_{k,u}|)^{-1}\left\{a^2(\eta_k) + \sum_{i=1}^{n-1}(\eta_{k,x_i})^2\right\}^{1/2}\right)|\eta_{k,u}|$$

$$\leq \sum_{k=1}^{2l} G\left((|y_{k,u}|)^{-1} \left\{ a^2(y_k) + \sum_{i=1}^{n-1} (y_{k,x_i})^2 \right\}^{1/2} \right) |y_{k,u}|. \tag{3.17}$$

Choosing

$$\lambda_j := \frac{|y_{j,u}|(y_j - y_1)}{\sum_{j=1}^{2l} |y_{j,u}|(y_j - y_1)}, \quad \mu_j := \frac{|y_{j,u}|(y_{2l} - y_j)}{\sum_{j=1}^{2l} |y_{j,u}|(y_{2l} - y_j)} \quad \text{and}$$

$$z_j := \left\{ a^2(y_j) + \sum_{i=1}^{n-1} (y_{j,x_i})^2 \right\}^{1/2} (|y_{j,u}|)^{-1}, \quad j = 1, \ldots, 2l,$$

the right-hand side of (3.17) becomes

$$\sum_{j=1}^{2l} \left(\frac{\sum_{k=1}^{2l} |y_{k,u}|(y_k - y_1)}{y_{2l} - y_1} \lambda_j F(z_j) + \frac{\sum_{k=1}^{2l} |y_{k,u}|(y_{2l} - y_k)}{y_{2l} - y_1} \mu_j F(z_j) \right) =: I.$$

Since G is convex we conclude from this

$$I \geq G\left(\sum_{j=1}^{2l} \lambda_j z_j \right) \frac{\sum_{k=1}^{2l} |y_{k,u}|(y_k - y_1)}{y_{2l} - y_1} + G\left(\sum_{j=1}^{2l} \mu_j z_j \right) \frac{\sum_{k=1}^{2l} |y_{k,u}|(y_{2l} - y_k)}{y_{2l} - y_1} =: I'.$$

Furthermore, from the monotonicity and convexity of the function $\varphi(\xi_1, \ldots, \xi_n) := \{\xi_1^2 + \cdots \xi_n^2\}^{1/2}$ we derive

$$\sum_{j=1}^{2l} \lambda_j z_j \geq \frac{\{(y_{2l} - y_1)^2 a^2(y_{2l}) + \sum_{i=1}^{n-1} [\sum_{j=1}^{2l} (-1)^j y_{j,x_i}(y_j - y_1)]^2\}^{1/2}}{\sum_{k=1}^{2l} |y_{k,u}|(y_k - y_1)}$$

and

$$\sum_{j=1}^{2l} \mu_j z_j \geq \frac{\{(y_{2l} - y_1)^2 a^2(y_1) + \sum_{i=1}^{n-1} [\sum_{j=1}^{2l} (-1)^j y_{j,x_i}(y_{2l} - y_j)^2]^2\}^{1/2}}{\sum_{k=1}^{2l} |y_{k,u}|(y_{2l} - y_k)}.$$

Together with the monotonocity of G this yields

$$\begin{aligned} I' \geq{} & \frac{\sum_{j=1}^{2l} |y_{j,u}|(y_j - y_1)}{y_{2l} - y_1} \\ & \times G\left(\frac{\{(y_{2l} - y_1)^2 a^2(y_{2l}) + \sum_{i=1}^{n-1} [\sum_{j=1}^{2l} (-1)^j y_{j,x_i}(y_j - y_1)]^2\}^{1/2}}{\sum_{j=1}^{2l} |y_{j,u}|(y_j - y_1)} \right) \\ & + \frac{\sum_{j=1}^{2l} |y_{j,u}|(y_{2l} - y_j)}{y_{2l} - y_1} \\ & \times G\left(\frac{\{(y_{2l} - y_1)^2 a^2(y_1) + \sum_{i=1}^{n-1} [\sum_{j=1}^{2l} (-1)^j y_{j,x_i}(y_{2l} - y_j)]^2\}^{1/2}}{\sum_{j=1}^{2l} |y_{j,u}|(y_{2l} - y_j)} \right). \end{aligned}$$

But in view of the identities (3.9)–(3.11) this last term is equal to the left-hand side of (3.17). Now from the inequalities (3.16) and (3.17) we obtain easily that the function $h(t) := I_t(x_0', u_0)$, $t \in [0, t_2)$, does not increase across the value $t = t_1$. Moreover, using the semigroup property we see that $h(t)$, $t \in [0, +\infty]$, is well-defined – with the except of a finite number of values t – and nonincreasing.

By Lemma 3.1 we can argue similarly for a.e. $(x',u) \in K$. This shows (3.13), and the theorem is proved. ■

A slight generalization of the previous Theorem 3.1 is the following:

COROLLARY 3.1

Let the functions $G(x',v,z)$, $a(x',y,v)$, $a_{ij}(x',v)$, $i,j = 1,\ldots,n-1$, *be continuou* $\forall (x,v,z) \in \mathbb{R}^n \times (\mathbb{R}_0^+)^2$. *Let* G *be nonnegative and convex in* z *with* $G(x',v,0) = ($ $\forall (x',v) \in \mathbb{R}^{n-1} \times \mathbb{R}_0^+$. *Further on let* a *be positive, even and convex in* y, *and let th matrix* (a_{ij}) *be positive definite. Finally let* u *be a good function. Then for every* $t \in$ $[0,+\infty]$ *we have*

$$\int_{\mathbb{R}^n} G\left(x', u, \left\{a^2 u_y^2 + \sum_{i,j=1}^{n-1} a_{ij} u_{x_i} u_{x_j}\right\}^{1/2}\right) \mathrm{d}x$$

$$\geq \int_{\mathbb{R}^n} G\left(x', u, \left\{\tilde{a}^2 (u_y^t)^2 + \sum_{i=1}^{n-1} \tilde{a}_{ij} u_{x_i}^t u_{x_j}^t\right\}^{1/2}\right) \mathrm{d}x. \qquad (3.18$$

(*For simplicity we wrote* $u = u(x)$, $u^t = u^t(x)$, $a = a(x,u(x))$, $\tilde{a} = a(x,u^t(x))$ *an* $a_{ij} = a_{ij}(x',u(x))$, $\tilde{a_{ij}} = a_{ij}(x',u^t(x))$, $i,j = 1,\ldots,n-1$, *in* (3.18).)

Proof. We fix an arbitrary point $(x_0',u_0) \in \mathbb{R}^{n-1} \times \mathbb{R}_0^+$. From the previous proof we se that it is sufficient to show the statements (3.16) and (3.17) at (x_0',u_0) – with the term containing partial derivatives in x_i, $(i = 1,\ldots,n-1)$, replaced by some correspondin quadratic forms. For an appropriate linear mapping $x' \mapsto \xi' \in \mathbb{R}^{n-1}$ we can achieve that th function $v(\xi,y) := u(x',y)$ satisfies

$$\sum_{i=1}^{n-1} \tilde{a}_{ij} u_{x_i} u_{x_j} = \sum_{i=1}^{n-1} \left(\frac{\partial v}{\partial \xi_i}\right)^2$$

at the point (x_0',u_0). Since, by the definition of the continuous symmetrization, $v^t(\xi',y) =$ $u^t(x',y)$, $(t \in [0,+\infty])$, (3.16) and (3.17) then follow as before. ■

Remark 3.2. (1) Integrals as in (3.12) and (3.18) with $G = G(z) = z^2$ and a some powe of y appear in variational problems for two-dimensional or axisymmetric flows (se [F, B1]). (2) In the case $t = +\infty$ (i.e. for Steiner symmetrization) the inequality (3.18) ca be proved in a simpler manner (see [B5]). Equation (3.18) seems to be the most genera Dirichlet-type inequality for Steiner symmetrization which appeared in the literatur Note that some similar inequalities with a radial weight function in the integrand are we known for the so-called starshaped rearrangements (see [BM, K1, 3, 4 and M]).

If $a = a_i \equiv 1$ and $G(z) = z^p$ for some $p \in (1,+\infty)$ in (3.18) then we are led to nor inequalities in $W^{1,p}$. For the proof we further need the following nice equivalenc principle for convex inequalities which was shown in ([ALT], Corollary 3.1).

Lemma 3.2. *Let* $u, v \in S_+(\mathbb{R}^n)$. *Then the following two properties* (i) *and* (ii) *a equivalent to each other,*

(i) $$\int_{\mathbb{R}^n} G(u)\mathrm{d}x \le \int_{\mathbb{R}^n} G(v)\mathrm{d}x, \quad \text{for every Young function } G.$$

(ii) $$\int_M u\,\mathrm{d}x \le \sup\left\{\int_N v\,\mathrm{d}x : |N| \le |M|,\ N \in \mathcal{M}(\mathbb{R}^n)\right\},$$

for every set $M \in \mathcal{M}(\mathbb{R}^n)$.

Theorem 3.2. *Let $u \in W^{1,p}(\mathbb{R}^n) \cap \mathcal{S}_+(\mathbb{R}^n)$ for some $p \in [1,+\infty]$. Then for every $t \in [0,+\infty]$ we have $u^t \in W^{1,p}(\mathbb{R}^n) \cap \mathcal{S}_+(\mathbb{R}^n)$ and*

$$\|\nabla u\|_p \ge \|\nabla u^t\|_p, \tag{3.19}$$

$$\left\|\frac{\partial u}{\partial x_i}\right\|_p \ge \left\|\frac{\partial u^t}{\partial x_i}\right\|_p, \qquad i = 1,\dots,n. \tag{3.20}$$

Proof. First observe that if $p \in [1,+\infty)$ and if u is good then (3.19) and (3.20) follow from Corollary 3.1. In the general case we will use various approximation arguments:

(1) Let $1 < p < +\infty$ and $u \in W_+^{1,p}(\mathbb{R}^n)$. We choose a sequence of good functions converging to u in $W^{1,p}(\mathbb{R}^n)$. From the equimeasurability it follows that $u^t, u_m^t \in L^p(\mathbb{R}^n)$, $m = 1,2,\dots$, and in view of the nonexpansivity, (Remark 2.6 (5)), we infer that $u_m^t \longrightarrow u^t$ in $L^p(\mathbb{R}^n)$. Further we have that $\|\nabla u_m^t\|_p \le \|\nabla u_m\|_p$ by (3.19), i.e. the functions u_m^t are uniformly bounded. Hence there is a subsequence $(u_{m'})^t$ which converges weakly in $W^{1,p}(\mathbb{R}^n)$ to some $v \in W^{1,p}(\mathbb{R}^n)$. This means that we have for every function $\varphi \in C_0^\infty(\mathbb{R}^n)$ and for every $i \in \{1,\dots,n\}$

$$-\int_{\mathbb{R}^n} u^t \frac{\partial\varphi}{\partial x_i}\mathrm{d}x \longleftarrow -\int_{\mathbb{R}^n} (u_{m'})^t \frac{\partial\varphi}{\partial x_i}\mathrm{d}x = \int_{\mathbb{R}^n} \varphi \frac{\partial(u_{m'})^t}{\partial x_i}\mathrm{d}x \longrightarrow \int_{\mathbb{R}^n} \varphi \frac{\partial v}{\partial x_i}\mathrm{d}x$$
$$\text{as } m' \to +\infty,$$

from which we can identify u^t as a function in $W^{1,p}(\mathbb{R}^n)$ with $\nabla u^t = \nabla v$. Since the norm in $W^{1,p}(\mathbb{R}^n)$ is weakly lower semicontinuous, we infer that

$$\|\nabla u^t\|_p \le \lim_{m'\to\infty}\inf \|\nabla (u_{m'})^t\|_p \le \lim_{m'\to\infty} \|\nabla u_{m'}\|_p = \|\nabla u\|_p. \tag{3.21}$$

Further we have $u_{x_i}^t \in L^p(\mathbb{R}^n)$, $i = 1,\dots,n$. Therefore an estimate for the partial derivatives $u_{x_i}^t$ analogous to (3.21) leads to the inequalities (3.20).

(Note that similar arguments can be found in ([K1], p. 23) and ([BZ], p. 159).)

(2) Let $u \in W^{1,\infty}(\mathbb{R}^n) \cap \mathcal{S}_+(\mathbb{R}^n)$. We introduce the cut-off functions u_c, $(c \ge 0)$, by

$$u_c := (u - c)_+ = \max\{u - c; 0\}. \tag{3.22}$$

Then

$$|\{u_c > 0\}| < +\infty \qquad \forall c > \inf u. \tag{3.23}$$

It follows that $u_c \in W^{1,p}(\mathbb{R}^n)$ for every $p \in [1,+\infty]$, and in view of $(u_c)^t = (u^t)_c$ we infer that

$$\|\nabla (u^t)_c\|_p \le \|\nabla u_c\|_p \qquad \forall p \in [1,+\infty). \tag{3.24}$$

Because of (3.23) we can pass to the limit $p \to +\infty$ in (3.24) to derive

$$\text{ess sup}\{|\nabla (u^t)_c(x) : x \in \mathbb{R}^n\} \le \text{ess sup}\{|\nabla u_c(x) : x \in \mathbb{R}^n\} \quad \forall c > \inf u.$$

Choosing $c \to 0$ and by taking into account that $\nabla u = 0$ a.e. on $\{u = \inf u\}$ and $\nabla u^t = 0$ a.e. on $\{u^t = \inf u\}$, (3.19) follows in the case $p = +\infty$. Analogous considerations lead to the inequalities (3.20) in the case $p = +\infty$.

(3) Let $u \in W_0^{1,1}(\mathbb{R}^n)$. We choose a sequence of Lipschitz continuous functions with compact support u_m converging to u in $W^{1,1}(\mathbb{R}^n)$. Then we have that $u_m^t \longrightarrow u^t$ in $L^1(\mathbb{R}^n)$. Furthermore, since $\|\partial u_m^t/\partial x_i\|_1 \leq \|\partial u_m/\partial x_i\|_1$ by (3.20), we see that the functions $(\partial u_m^t)/(\partial x_i)$ are uniformly bounded in $L^1(\mathbb{R}^n)$, $i = 1, \dots, n$, and for every Young function G we have

$$\int_{\mathbb{R}^n} G\left(\left|\frac{\partial u_m^t(x)}{\partial x_i}\right|\right) dx \leq \int_{\mathbb{R}^n} G\left(\left|\frac{\partial u_m(x)}{\partial x_i}\right|\right) dx,$$
$$m = 1, 2, \dots, \quad i = 1, \dots, n. \tag{3.25}$$

From Lemma 3.2 we infer that for every set $M \in \mathcal{M}(\mathbb{R}^n)$

$$\int_M \left|\frac{\partial u_m^t(x)}{\partial x_i}\right| dx \leq \sup\left\{\int_N \left|\frac{\partial u_m(x)}{\partial x_i}\right| dx : |N| \leq |M|,\ N \in \mathcal{M}(\mathbb{R}^n)\right\},$$
$$m = 1, 2, \dots, \quad i = 1, \dots, n. \tag{3.26}$$

Now assume for a moment that for every $i \in \{1, \dots, n\}$

$$\sup\left\{\int_{E_k} \left|\frac{\partial u_m^t(x)}{\partial x_i}\right| dx : m \in \mathbb{N}\right\} \longrightarrow 0 \qquad \text{as } k \to +\infty, \tag{3.27}$$

for every sequence $\{E_k\} \subset \mathcal{M}(\mathbb{R}^n)$ with $|E_k| \longrightarrow 0$.

From a well known weak compactness principle of sequences in $L^1(\mathbb{R}^n)$, (see e.g. [Alt], p. 199), we infer that there are subsequences $(\partial u_{m'}^t)/(\partial x_i)$ which converge weakly in $L^1(\mathbb{R}^n)$ to functions $v_i \in L^1(\mathbb{R}^n)$, respectively, $i = 1, \dots, n$. By proceeding as in part (1) of the proof one obtains then (3.19) and (3.20). Thus it remains to show (3.27).

Suppose that (3.27) is not true for some $i \in \{1, \dots, n\}$. In view of (3.26) there is a number $\delta > 0$ and sequences $\{m_k\} \subset \mathbb{N}$ and $\{E_k\} \subset \mathcal{M}(\mathbb{R}^n)$ such that $|E_k| \longrightarrow 0$ as $k \to +\infty$ and

$$\sup\left\{\int_N \left|\frac{\partial u_{m_k}(x)}{\partial x_i}\right| dx : |N| \leq |E_k|, N \in \mathcal{M}(\mathbb{R}^n)\right\} \geq \delta. \tag{3.28}$$

Therefore we can find a sequence $\{N_k\} \subset \mathcal{M}(\mathbb{R}^n)$ with $|N_k| \leq |E_k|$, $k = 1, 2, \dots$, such that

$$\int_{N_k} \left|\frac{\partial u_{m_k}(x)}{\partial x_i}\right| dx \geq \frac{\delta}{2}. \tag{3.29}$$

There are possible two cases:

(a) The sequence $\{m_k\}$ is unbounded. We choose a subsequence $\{k'\}$ with $m_{k'} \to +\infty$ as $k' \to +\infty$. From (3.29) we have that

$$\int_{N'_k} \left|\frac{\partial u(x)}{\partial x_i}\right| dx \geq \frac{\delta}{4} \qquad \text{for } k' \text{ large enough},$$

which is impossible since $u \in W^{1,1}(\mathbb{R}^n)$.

(b) The sequence $\{m_k\}$ is bounded. Then by passing to a subsequence $\{k'\}$ with $m_{k'} = m$ ($=$ const) in (3.29) we derive a contradiction to $u_m \in W^{1,1}(\mathbb{R}^n)$. ■

Remark 3.3. There exists a simple alternative proof of Theorem 3.2 by means of an approximation via convolution-type inequalities (see [B4]). Note that this method of proof was developped by Baernstein [Ba] for various types of rearrangements. Unfortunately, this idea seems not applicable in the case of the general inequalities (3.12) and (3.18).

It is easy to obtain an analogue of Theorem 3.2 for functions in the Sobolev spaces $W^{1,p}_{0+}(\Omega)$, (Ω open).

COROLLARY 3.2

Let Ω be an open set and let $u \in W^{1,p}_{0+}(\Omega)$ for some $p \in (1, +\infty)$. Then for every $t \in [0, +\infty]$ we have $u^t \in W^{1,p}_{0+}(\Omega^t)$ and (3.19), (3.20) *hold.*

Proof. Equations (3.19) and (3.20) follow from Theorem 3.2 by extending u and u^t by zero outside Ω and Ω^t, respectively. Thus it remains to show that $u^t \in W^{1,p}_0(\Omega^t)$. If $u \in C^{0,1}_{0+}(\Omega)$ it follows by Remark 2.6 (8) that $u^t \in C^{0,1}_{0+}(\Omega^t)$. In the general case we choose a sequence u_m of functions in $C^{0,1}_{0+}(\Omega)$ which converges to u in $W^{1,p}_0(\Omega)$. Then $(u_m)^t \to u^t$ in $L^p(\Omega^t)$. By (3.19) the functions $(u_m)^t$ are equibounded in $W^{1,p}_0(\Omega^t)$. Therefore there is a function $v \in W^{1,p}_0(\Omega^t)$ and a subsequence $(u_{m'})^t$ which converges to v weakly in $W^{1,p}_0(\Omega^t)$. This means that for every $\varphi \in C^\infty_0(\Omega^t)$ and for every $i \in \{1, \dots, n\}$

$$\int_{\Omega^t} \varphi v_{x_i}\, \mathrm{d}x \longleftarrow \int_{\Omega^t} \varphi \frac{\partial (u_{m'})^t}{\partial x_i}\, \mathrm{d}x = -\int_{\Omega^t} \varphi_{x_i} (u_{m'})^t\, \mathrm{d}x \longrightarrow -\int_{\Omega^t} \varphi_{x_i} u^t \mathrm{d}x \qquad \text{as } m' \to +\infty,$$

that is $v = u^t$. The corollary is proved. ■

The following property is useful for approximations of the symmetrized functions. It can be proved by arguing as in part (1) of the proof of Theorem 3.2.

Lemma 3.3. *Let $u, u_m \in W^{1,p}_+(\mathbb{R}^n)$, $m = 1, 2, \dots$, for some $p \in (1, +\infty)$ and*

$$u_m \longrightarrow u \qquad \text{in } W^{1,p}(\mathbb{R}^n) \qquad \text{as } m \to +\infty. \tag{3.30}$$

Then for every $t \in [0, +\infty]$

$$u^t_m \rightharpoonup u^t \qquad \text{weakly in } W^{1,p}(\mathbb{R}^n) \qquad \text{as } m \to +\infty. \tag{3.31}$$

Open problem 3.1. Let u, u_m be as in Lemma 3.3. Is it then true that for every $t \in [0, +\infty]$

$$u^t_m \longrightarrow u^t \qquad \text{in } W^{1,p}(\mathbb{R}^n) \qquad \text{as } m \to +\infty\,? \tag{3.32}$$

This conjecture was shown in the case $t = +\infty$ (i.e. for the Steiner symmetrization) by Burchard [Bu] (see also [C] for an earlier proof in the particular case that $n = 1$ and $t = +\infty$).

It is worth to mention that, if $n \geq 2$, then a conclusion analogous to (3.32) does not hold for the Schwarz symmetrizations of u, u_m, $m = 1, 2, \dots$, (see [AL]).

It is possible to extend Corollary 3.1 to Sobolev functions.

COROLLARY 3.3

Let Ω be an open set with $\Omega = \Omega^*$ and let $u \in W^{1,p}_{0+}(\Omega)$ for some $p \in [1, +\infty)$. Further let G, a, a_{ij}, $i, j = 1, \dots, n-1$, be as in Corollary 3.1, *and suppose that for some numbers*

$C > c > 0$

$$c \leq a(x, v) \leq C, \qquad c\sum_{i=1}^{n-1}\xi_i^2 \leq \sum_{i,j=1}^{n-1} a_{ij}(x', v)\xi_i\xi_j \leq C\sum_{i=1}^{n-1}\xi_i^2$$
$$\forall(\xi_1, \dots, \xi_{n-1}) \in \mathbb{R}^{n-1} \quad \mathit{and} \quad \forall(x, v, z) \in \mathbb{R}^n \times (\mathbb{R}_0^+)^2. \qquad (3.33$$

Finally suppose that

$$G(x', v, z) \leq \begin{cases} Cz^p & \mathit{if}\ |\Omega| = +\infty \\ C(z + z^p) & \mathit{if}\ |\Omega| < +\infty \end{cases}. \qquad (3.34$$

Then (3.18) *holds.*

Proof. We choose a sequence of good functions $\{u_m\}$ such that

$$\begin{aligned} &u_m \longrightarrow u \text{ in } W^{1,p}(\Omega) \quad \text{and} \\ &u_m \longrightarrow u \\ &\nabla u_m \longrightarrow \nabla u \end{aligned} \qquad \text{a.e. in } \Omega. \qquad (3.35$$

Let $J(u)$ denote the integral functional on the left-hand side of (3.18). By (3.33) we ha

$$J(u_m) \leq \begin{cases} C\|\nabla u_m\|_p^p & \text{if } |\Omega| = +\infty \\ C(\|\nabla u_m\|_1 + \|\nabla u_m\|_p^p) & \text{if } |\Omega| < +\infty \end{cases}, \qquad (3.36$$

and the same inequality holds for u_m replaced by u. In view of (3.34), (3.35) we can app Lebesgue's convergence theorem to infer that $\lim_{m\to+\infty} J(u_m) = J(u)$.

Let $t \in [0, +\infty]$. Since the functions $(u_m)^t$ are equibounded in $W^{1,p}(\Omega)$ we can choo a subsequence $\{(u_{m'})^t\}$ which converges to u^t weakly in $W^{1,p}(\Omega)$. In view of the we lower semicontinuity of the functional J this finally gives

$$J(u^t) \leq \liminf_{m'\to+\infty} J((u_{m'})^t) \leq \lim_{m'\to+\infty} J(u_{m'}) = J(u).$$

Remark 3.4. (1) The inequalities (3.19) find their analogy in inequalities for the norm the space of functions with bounded variation in $\mathbf{R}^n$ (see [B4]). A consequence of this that the perimeter of Caccioppoli sets decreases under continuous symmetrization. (With regard to some 'nice' properties of the continuous symmetrization – and particular to the basic fact that the Lipschitz continuity of functions is preserved und continuous rearrangement – our restriction to the Sobolev spaces $W^{1,p}(\mathbb{R}^n)$ is r forcible. The general Dirichlet-type inequality (3.18), for instance, is also satisfied f functions lying in a suitable Orlicz space.

4. Continuity in *t*

In this section we are interested in continuity properties of the mapping

$$t \longmapsto u^t$$

in the spaces $L^p(\mathbb{R}^n)$, $W^{1,p}(\mathbb{R}^n)$, $1 \leq p < +\infty$, and $BV(\mathbb{R}^n)$.

Lemma 4.1. *Let $u \in L^p_+(\mathbb{R}^n)$ for some $p \in [1, +\infty)$ and let $\{t_m\}$ be a nonnegative sequence converging to some number $t \in [0, +\infty]$. Then*

$$u^{t_m} \longrightarrow u^t \quad \text{in } L^p(\mathbb{R}^n) \quad \text{as } m \to +\infty. \tag{4.1}$$

Proof. The proof is in two steps:

(1) Let u be a step function of the following form,

$$u = \varepsilon \sum_{i=1}^{k} \chi(M_i), \tag{4.2}$$

where $M_1 \supset \cdots \supset M_k$, $M_i \in \mathcal{M}(\mathbb{R}^n)$, $i = 1, \ldots, k$, $\varepsilon > 0$.
Then

$$u^t = \varepsilon \sum_{i=1}^{k} \chi(M_i^t), \quad u^{t_m} = \varepsilon \sum_{i=1}^{k} \chi(M_i^{t_m}), \qquad m = 1, 2, \ldots.$$

In view of (2.14) we have

$$\begin{aligned}\|u^{t_m} - u^t\|_p &= \varepsilon \left\| \sum_{i=1}^{k} (\chi(M_i^{t_m}) - \chi(M_i^t)) \right\|_p \\ &\le \varepsilon \sum_{i=1}^{k} |M_i^{t_m} \Delta M_i^t| \longrightarrow 0 \qquad \text{as } m \to +\infty.\end{aligned}$$

(2) Let $u \in L^p_+(\mathbb{R}^n)$ and $\varepsilon > 0$. We choose a sequence of step functions $\{u_k\}$ which converges to u in $L^p(\mathbb{R}^n)$. We may take k large enough such that $\|u_k - u\|_p < \varepsilon/3$ and then m large enough to ensure that $\|u_k^{|t_m - t|} - u_k\|_p < \varepsilon/3$. In view of the nonexpansivity (2.35) we derive

$$\begin{aligned}\|u^{t_m} - u^t\|_p &\le \|u^{t_m} - u_k^{t_m}\|_p + \|u_k^{t_m} - u_k^t\|_p + \|u_k^t - u^t\|_p \\ &\le 2\|u_k - u\|_p + \|u_k^{|t_m - t|} - u_k\|_p < \varepsilon,\end{aligned}$$

and the assertion follows. ∎

Lemma 4.2. *Let $u \in L^p(\mathbb{R}^n)$ and $u \neq u^*$. Then there are constants $c > 0$ and $t_0 > 0$ such that:*

$$\|u^t - u\|_p \ge ct \qquad \forall t \in [0, t_0]. \tag{4.3}$$

Proof. By Lemma 4.1 we can find numbers $t_0 > 0$ and $\delta > 0$ such that $\|u^t - u\|_p \ge \delta$ $\forall t \in [t_0, +\infty]$. If $t \in [0, t_0]$ we find some number $N \in \mathbb{N}$ satisfying $t_0 \le Nt \le 2t_0$. Then by the nonexpansivity we derive

$$\|u^{Nt} - u\|_p \le \sum_{k=0}^{N-1} \|u^{(k+1)t} - u^{kt}\|_p \le N\|u^t - u\|_p,$$

which means that

$$\|u^t - u\|_p \ge \frac{\delta}{N} \ge \frac{\delta}{2t_0} t. \qquad ∎$$

Theorem 4.1. *Continuity from the right of the mapping* $t \longmapsto u^t$: *Let* $t_m \searrow 0$ *and* $u \in W_+^{1,p}(\mathbb{R}^n)$ *for some* $p \in [1, +\infty)$. *Then*

$$u^{t_m} \longrightarrow u \qquad \text{in } W^{1,p}(\mathbb{R}^n) \qquad \text{as } m \to +\infty. \tag{4.4}$$

Proof. Let $i \in \{1, \ldots, n\}$. We split into two cases:

(1) $p > 1$. From Theorem 3.2 we infer that the sequence $\|u_{x_i}^{t_m}\|_p$ is monotonically increasing and

$$\lim_{m\to\infty} \|u_{x_i}^{t_m}\|_p \leq \|u_{x_i}\|_p. \tag{4.5}$$

Furthermore, the sequence $\{u^{t_m}\}$ converges to u in $L^p(\mathbb{R}^n)$ by Lemma 4.1. It follows that for every $\varphi \in C_0^\infty(\mathbb{R}^n)$

$$-\int_{\mathbb{R}^n} \varphi \frac{\partial u^{t_m}}{\partial x_i} \mathrm{d}x = \int_{\mathbb{R}^n} u^{t_m} \varphi_{x_i} \, \mathrm{d}x \longrightarrow \int_{\mathbb{R}^n} u \varphi_{x_i} \mathrm{d}x = -\int_{\mathbb{R}^n} \varphi u_{x_i} \mathrm{d}x,$$

that is

$$\frac{\partial u^{t_m}}{\partial x_i} \rightharpoonup \frac{\partial u}{\partial x_i} \quad \text{weakly in } L^p(\mathbb{R}^n) \quad \text{as } m \to +\infty. \tag{4.6}$$

Since the spaces $L^p(\mathbb{R}^n)$ are uniformly convex if $1 < p < +\infty$, (4.5) and (4.6) imply that

$$\frac{\partial u^{t_m}}{\partial x_i} \longrightarrow \frac{\partial u}{\partial x_i} \qquad \text{strongly in } L^p(\mathbb{R}^n), \qquad i = 1, \ldots, n.$$

(2) $p = 1$. As in part (1) we can derive (4.5) and (4.6). We set $v_m := (u^{t_m})_{x_i}$ and $v := (u^t)_{x_i}$. Since the function $G(z) := \sqrt{1 + z^2} - 1$ is continuous and convex with $G(z) \leq z$ we conclude from Corollary 3.4 and the weak lower semi-continuity of the integral that

$$\lim_{m\to\infty} \int_{\mathbb{R}^n} (\sqrt{1 + v_m^2} - 1) \mathrm{d}x = \int_{\mathbb{R}^n} (\sqrt{1 + v^2} - 1) \mathrm{d}x.$$

Further we obtain by Taylor's theorem

$$\int_{\mathbb{R}^n} (\sqrt{1 + v_m^2} - 1) \mathrm{d}x \geq \int_{\mathbb{R}^n} (\sqrt{1 + v^2} - 1) \mathrm{d}x + \int_{\mathbb{R}^n} \frac{v}{\sqrt{1 + v^2}} (v_m - v) \mathrm{d}x$$
$$+ \frac{1}{2} \int_{\mathbb{R}^n} \frac{(v_m - v)^2}{(1 + \max\{v^2; v_m^2\})^{3/2}} \mathrm{d}x.$$

By passing to the limit $m \to +\infty$ this leads to

$$\lim_{m\to\infty} \int_{\mathbb{R}^n} \frac{(v_m - v)^2}{(1 + \max\{v^2; v_m^2\})^{3/2}} \mathrm{d}x = 0.$$

In particular this means that

$$\lim_{m\to\infty} \int_{\{|v_m|, |v| \leq k\}} |v_m - v| \mathrm{d}x = 0 \qquad \forall k > 0. \tag{4.7}$$

Since $v_m \rightharpoonup v$ weakly in $L^p(\mathbb{R}^n)$ we have also

$$\lim_{k\to+\infty} \int_{\{|v_m| > k\}} |v_m| \mathrm{d}x = 0 \qquad \text{uniformly } \forall m \in \mathbb{N}. \tag{4.8}$$

Now the assertion follows easily from (4.7), (4.8) and from the inequalities

$$\|v_m - v\|_1 \leq \int_{\{|v_m|,\, |v|\leq k\}} |v_m - v|dx + 2\int_{\{|v_m|>k\}} |v_m|dx + 2\int_{\{|v|>k\}} |v|dx \qquad \forall k > 0. \qquad \blacksquare$$

Remark 4.1. Simple examples of piecewise linear functions show that Theorem 4.1 does not hold in the case $p = +\infty$.

Open problem 4.1. It would be interesting to find out whether the mapping

$$t \longmapsto u^t$$

is also continuous from the left in $W_+^{1,p}(\mathbb{R}^n)$, $1 \leq p < +\infty$.

Next we want to estimate $\|u^t - u\|_p$ from above for functions in Sobolev spaces.

Lemma 4.3. *Let u be a good function. Then the limit function*

$$U(x) := \lim_{t\to 0} \frac{1}{t}(u^t(x) - u(x)) \tag{4.9}$$

exists a.e. *Moreover if u is differentiable in (x', y_i), $i = 1, 2$, and*

$$y_1 < y_2, \quad u(x', y_1) = u(x', y_2) < u(x', z) \qquad \forall z \in (y_1, y_2), \tag{4.10}$$

then

$$U(x', y_i) = \frac{y_1 + y_2}{2} u_y(x', y_i), \qquad i = 1, 2. \tag{4.11}$$

Proof. For almost every $x^1 = (x', y_1)$ with $u_y(x^1) > 0$ we can find a point $x^2 = (x', y_2)$ such that $u_y(x^2) < 0$ and such that (4.10) is satisfied. We fix two points x^1 and x^2 with these properties. Let $y_i = y_i(x', u)$ be the local inverse function of u in the neighborhood of $x^i, i = 1, 2$, respectively. Then for small enough $t > 0$, the function u^t can be represented by corresponding local inverse functions $y_i^t = y_i^t(x', u)$, $i = 1, 2$, which are given by the formulas (2.10). In other words, we have for small $t > 0$,

$$u^t(x', y_i^t(x', u)) = u(x', y_i), \quad i = 1, 2.$$

Differentiating this we obtain, using (2.10),

$$\left.\frac{\partial u^t(x', y_i)}{\partial t}\right|_{t=0} = -\frac{\partial u^t(x', y_i)}{\partial y} \cdot \left.\frac{\partial y_i^t}{\partial t}\right|_{t=0} = \frac{y_1 + y_2}{2} u_y(x', y_i), \qquad i = 1, 2.$$

Reversely, for almost every $x^2 = (x', y_2)$ with $u_y(x^2) < 0$ we can find a point $x^1 = (x', y_1)$ such that $u_y(x^1) > 0$ and such that (4.10) is satisfied, and we conclude as before. $\blacksquare$

Theorem 4.2. *Let $u \in W_{0+}^{1,p}(B_R)$ for some $p \in [1, +\infty]$ and $R > 0$. Then*

$$\|u^t - u\|_p \leq tR\|u_y\|_p \qquad \forall t \in [0, +\infty]. \tag{4.12}$$

Proof. Let u be Lipschitz continuous and let $x_0 = (x_0', y_0) \in B_R$. We set

$$u_1(y) := \max\{0; u(x_0) - \|u_y\|_\infty |y - y_0|\} \qquad \text{and}$$
$$u_2(y) := \max\{0; \min\{u(x_0) + \|u_y\|_\infty |y - y_0|;\ \|u_y\|_\infty (R - |y|)\}\} \qquad \forall y \in \mathbb{R}.$$

Clearly we have $u_1(y) \le u(x_0) \le u_2(y)$ and $u_1(y) \le u(x_0', y) \le u_2(y)$ $\forall y \in \mathbb{R}$. Let u_i^t deno the (one-dimensional !) continuous symmetrization of the function u_i, $i = 1, 2$, respe tively. We obtain by monotonicity

$$u_1^t(y) \le u(x_0) \le u_2^t(y) \quad \text{and} \quad u_1^t(y) \le u^t(x_0', y) \le u_2^t(y) \quad \forall y \in \mathbb{R}. \tag{4.13}$$

Furthermore, a simple computation shows that

$$\max\{u(x_0) - u_1^t(y_0);\ u_2^t(y_0) - u(x_0)\} \le tR\|u_y\|_\infty .$$

Together with (4.13) this yields $|u^t(x_0) - u(x_0)| \le tR\|u_y\|_\infty$, which proves (4.12) in th case $p = +\infty$.

Next let $1 \le p < +\infty$. First assume that u is a good function. From (4.11) we obta $|U(x)| \le R|u_y(x)|$ for a.e. $x \in B_R$. After an integration over B_R this yields

$$\|U\|_p \le R\|u_y\|_p . \tag{4.14}$$

The functions $(1/t)(u^t - u)$ are equibounded in $L^\infty(B_R)$ by (4.12) and converge to U a. in B_R. By applying Lebesgue's convergence theorem we infer that

$$\frac{u^t - u}{t} \longrightarrow U \quad \text{in } L^p(B_R) \qquad \text{as } t \to 0. \tag{4.15}$$

Further we derive from the nonexpansivity

$$\|u^t - u\|_p \le \sum_{k=0}^{N-1} \|u^{(k+1)t/N} - u^{kt/N}\|_p \le N\|u^{t/N} - u\|_p \qquad \forall N \in \mathbb{N}.$$

By passing to the limit $N \to +\infty$, this yields $\|u^t - u\|_p \le t\|U\|_p$ in view of (4.15). No the assertion follows from (4.14).

In the general case we choose a sequence $\{u_m\}$ of good functions converging to u $W^{1,p}(B_R)$ and compute

$$\|u^t - u\|_p \le tR\|(u_m)_y\|_p + \|u_m - u\|_p + \|u_m^t - u^t\|_p \longrightarrow tR\|u_y\|_p \quad \text{as } m \to +$$

The theorem is proved.

Now we prove an analogue of Lemma 4.1 for functions in $C(\mathbb{R}^n) \cap \mathcal{S}_+(\mathbb{R}^n)$. Note th Lemma 4.5 below generalizes a part of Theorem 7 in [B2].

Lemma 4.4. *Let* $u \in C(\mathbb{R}^n) \cap \mathcal{S}_+(\mathbb{R}^n)$ *and* $t, \{t_m\}$ *as in Lemma* 4.1. *Then*

$$\|u^{t_m} - u^t\|_\infty \longrightarrow 0 \qquad \text{as } m \to +\infty. \tag{4.16}$$

Proof. If $u \in C_{0+}^{0,1}(B_R)$ for some $R > 0$, we obtain from Theorem 4.2 and (4.12) t estimate $\|u^{t_m} - u^t\|_\infty \le R\ |t_m - t|\|u_y\|_\infty$. In the general case let $\varepsilon > 0$. We choose nonnegative Lipschitz function v with compact support such that $\|u - v\|_\infty < \varepsilon/3$. setting $R :=$ diam (supp v) we find some $m_0 \in \mathbb{N}$ such that $|t_m - t| < \varepsilon(3R\|v_y\|_\infty)$ $\forall m \ge m_0$. By the nonexpansivity we have for every $m \ge m_0$

$$\begin{aligned} \|u^{t_m} - u^t\|_\infty &\le \|u^{t_m} - v^{t_m}\|_\infty + \|v^{t_m} - v^t\|_\infty + \|v^t - u^t\|_\infty \\ &\le 2\|u - v\|_\infty + \|v^{t_m} - v^t\|_\infty < \varepsilon . \end{aligned}$$

The lemma is proved.

Formula (4.11) shows that the difference $|u^t(x) - u(x)|$ is proportional to $|x|$. Therefore it is not easy to derive estimates like Theorem 4.2 for functions which do not have bounded support. However this is possible if u satisfies some decaying properties near infinity. We study a typical situation.

Theorem 4.3. *Let $u \in W^{1,p}(\mathbb{R}^n) \cap \mathcal{S}_+(\mathbb{R}^n)$ for some $p \in [1, +\infty]$. Let $\varphi : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ be continuous and decreasing, and suppose that φ satisfies*

$$\int_0^{+\infty} r^{(n/p)-1}\varphi(r)\mathrm{d}r < +\infty. \tag{4.17}$$

Further suppose that u satisfies $(\alpha \in (0,1), R, d > 0)$

$$u(x) \geq \varphi(R/\alpha) \qquad \text{if } |x| \leq R \tag{4.18}$$

and

$$\left.\begin{array}{r}\varphi(|x|/\alpha) \leq u(x) \leq \varphi(|x|) \\ |\nabla u(x)| \leq d|x|^{-1}\varphi(|x|)\end{array}\right\} \quad \text{if } |x| \geq R. \tag{4.19}$$

Then there is some constant $C > 0$, depending only on φ, α, R, d and $\|u_y\|_p$, such that

$$\|u^t - u\|_p \leq Ct \qquad \forall t \in [0, +\infty]. \tag{4.20}$$

Furthermore, (4.18) *and the first inequality in* (4.19) *remain valid for u replaced by u^t for every $t \in [0, +\infty]$.*

Proof. Let $t \in [0, +\infty]$. The idea of the proof consists in combining the formulas (4.12) and (4.19) with a 'layer cake' argument.

We introduce the functions

$$u_0 := \max\{u - \varphi(R); 0\} \quad \text{and}$$

$$u_i := \begin{cases} \varphi(2^{i-1}R) - \varphi(2^iR) & \text{if } u > \varphi(2^{i-1}R) \\ u - \varphi(2^iR) & \text{if } \varphi(2^iR) < u \leq \varphi(2^{i-1}R), \quad i = 1, 2, \ldots. \\ 0 & \text{if } u \leq \varphi(2^iR) \end{cases}$$

We have

$$u = \sum_{i=0}^{+\infty} u_i, \tag{4.21}$$

and in view of Definition 2.4

$$u^t = \sum_{i=0}^{+\infty} (u_i)^t. \tag{4.22}$$

From the assumptions we see that supp $u_0 \subset \overline{B_R}$. By applying Theorem 4.2 this leads to

$$\|(u_0)^t - u_0\|_p \leq tR\left\|\frac{\partial u_0}{\partial y}\right\|_p \leq tR\|u_y\|_p. \tag{4.23}$$

Further, if $i \in \mathbb{N}$, we see from (4.19) that supp $u_i \subset \overline{B_{2^iR}}$ and $\nabla u_i(x) = 0$ in $B_{2^{i-1}\alpha R}$. By using (4.12) and (4.18) again we obtain

$$\begin{aligned}\|(u_i)^t - u_i\|_p &\le t\,2^iR\left\|\frac{\partial u_i}{\partial y}\right\|_p \le t\,2^iR\|\nabla u_i\|_\infty |\mathrm{supp}|\nabla u_i||^{1/p}\\ &\le t\,2^iRd(2^{i-1}\alpha R)^{-1}\varphi(2^{i-1}\alpha R)(\omega_n)^{1/p}((2^iR)^n - (2^{i-1}\alpha R)^n)^{1/p}_.\\ &\le t\,(\omega_n)^{1/p}d\alpha^{-1}(2^iR)^{n/p}\varphi(2^{i-1}\alpha R) \equiv tC_i. \end{aligned} \tag{4.24}$$

In view of (4.17) we derive easily

$$\sum_{i=1}^{+\infty} C_i < +\infty. \tag{4.25}$$

Finally, by using (4.21)–(4.25) we obtain

$$\|u^t - u\|_p \le \sum_{i=0}^{+\infty} \|(u_i)^t - u_i\|_p \le t\left(R\|u_y\|_p + \sum_{i=1}^{+\infty} C_i\right) \equiv tC,$$

and (4.20) follows. Now set

$$\varphi_1(x) := \begin{cases} \varphi(|x|/\alpha) & \text{if } |x| > R\\ \varphi(R/\alpha) & \text{if } |x| \le R \end{cases} \quad \text{and}$$

$$\varphi_2(x) := \begin{cases} \varphi(|x|) & \text{if } |x| > R\\ +\infty & \text{if } |x| \le R \end{cases}.$$

We have $\varphi_1 \le u \le \varphi_2$, by (4.18), (4.19), and clearly $\varphi_i = (\varphi_i)^*$, $i = 1, 2$. By the monotonicity of continuous symmetrization this means that $\varphi_1 = (\varphi_1)^t \le u^t \le (\varphi_2)^t = \varphi_2$ $\forall t \in [0, +\infty]$, which proves the second assertion of the lemma. ■

Remark 4.2. It is easy to verify that the function u in the cases (1) and (2) below satisfies the assumptions of Theorem 4.4 ($R > 1, \delta, c_1, c_2, c_3, \gamma, \lambda > 0, \sigma, \tau \in \mathbb{R}$).

(1) $u \in W_+^{1,p}(\mathbb{R}^n)$ for some $p \in [1, +\infty]$ and u satisfies

$$u(x) \ge \delta \quad \text{if } |x| \le R \tag{4.26}$$

and one of the following conditions (i) or (ii)

(i) $$\gamma > (n/p) \tag{4.27}$$

and

$$\left.\begin{aligned} c_1|x|^{-\gamma}(\log|x|)^{-\sigma} \le u(x) \le c_2|x|^{-\gamma}(\log|x|)^{-\sigma}\\ |\nabla u(x)| \le c_3|x|^{-\gamma-1}(\log|x|))^{-\sigma} \end{aligned}\right\} \text{ if } |x| \ge R,$$

(ii) $$\left.\begin{aligned} c_1e^{-\lambda|x|}|x|^\tau \le u(x) \le c_2e^{-\lambda|x|}|x|^\tau\\ |\nabla u(x)| \le c_3e^{-\lambda|x|}|x|^\tau \end{aligned}\right\} \text{ if } |x| \ge R. \tag{4.28}$$

(2) $u \in W^{1,\infty}(\mathbb{R}^n) \cap S_+(\mathbb{R}^n)$, $\sigma > 0$ and u satisfies (4.26) and

$$\left.\begin{aligned} (\log(c_1|x|))^{-\sigma} \le u(x) \le (\log(c_2|x|))^{-\sigma}\\ |\nabla u(x)| \le c_3|x|^{-1}(\log(c_2|x|))^{-\sigma-1} \end{aligned}\right\} \text{ if } |x| \ge R. \tag{4.29}$$

5. More estimates

In this section we intend to show estimates of the form

$$\liminf_{t\searrow 0}\frac{1}{t}\int_\Omega f(x,u)(u^t-u)\mathrm{d}x \geq 0, \qquad u\in W_+^{1,p}(\Omega), \tag{5.1}$$

for suitable functions f and domains Ω in $\mathbb{R}^n$. At this stage it is worth to sketch a proof of (5.1) in a special case.

Let $f=f(u)$ be continuous and $F(u)=\int_0^u f(v)\mathrm{d}v$, and let Ω be a bounded domain with $\Omega=\Omega^*$. The equimeasurability of u and u^t yields $\int F(u)=\int F(u^t)$. Furthermore, $f(u)\,(u^t-u)$ represents the first summand in the (formal) Taylor expansion of the difference $F(u^t)-F(u)$ into powers of (u^t-u). Heuristically this means that $\int f(u)(u^t-u)$ is small in $\|u^t-u\|_p$. Applying Lemma 4.2 and Theorem 4.2 this finally gives

$$\int_\Omega f(u)(u^t-u)\mathrm{d}x = o(t) \quad \text{as } t\searrow 0.$$

The next Theorem 5.1 can be seen in a certain sense as a generalization of the Hardy–Littlewood inequality (2.36).

Theorem 5.1. *Let Ω be an open set with $\Omega=\Omega^*$. Let $u\in L_+^p(\mathbb{R}^n)$ for some $p\in[1,+\infty)$, and suppose that u vanishes outside Ω. Furthermore, let $F=F(x,v)$ measurable on $\Omega\times[0,\sup u]$, continuous in v and satisfies*

$$\begin{aligned}&F(x,0)=0 \qquad \forall x\in\Omega,\\ &|F(x,v)|\leq A(x)B(x',v) \qquad \forall (x,v)\in\Omega\times[0,\sup u],\end{aligned} \tag{5.2}$$

where $B(x',v)$ is nonnegative, measurable in x' and nondecreasing and continuous in v, and

$$A\in L_+^{1/(1-\alpha)}(\Omega), B(\cdot,u(\cdot))\in L^{1/\alpha}(\Omega)$$

for some $\alpha\in[0,1]$.

Finally, suppose that for every $s>0$ the function

$$\varphi_s(x,v) := F(x,v+s)-F(x,v) \tag{5.3}$$

is symmetrically nonincreasing in y. Then

$$\int_\Omega F(x,u)\mathrm{d}x \leq \int_\Omega F(x,u^t)\mathrm{d}x. \tag{5.4}$$

Remark 5.1. If F is differentiable in v, then the condition (5.3) means that $(\partial F/\partial v)(x,v)$ is symmetrically nonincreasing in y.

Proof of Theorem 5.1. The proof is in two steps:

(1) First assume that u is a step function of the form (4.2). We compute

$$\int_\Omega F(x,u)\mathrm{d}x = \sum_{i=1}^k \int_{M_i}(F(x,\varepsilon i)-F(x,\varepsilon(i-1)))\mathrm{d}x \quad \text{and}$$

$$\int_\Omega F(x,u^t)\mathrm{d}x = \sum_{i=1}^k \int_{M_i^t}(F(x,\varepsilon i)-F(x,\varepsilon(i-1)))\mathrm{d}x. \tag{5.5}$$

Note that the integrals on the right-hand sides in (5.5) converge by (5.2). By (5.3) the functions $\varphi_i(x) := F(x, \varepsilon i) - F(x, \varepsilon(i-1))$ are symmetrically nonincreasing in y. We claim that

$$\int_{M_i} \varphi_i(x)\mathrm{d}x \leq \int_{M_i^t} \varphi_i(x)\mathrm{d}x, \quad i = 1, \dots, k. \tag{5.6}$$

To this end it is sufficient to prove that

$$\int_{M_i(x')} \varphi_i(x', y)\mathrm{d}y \leq \int_{(M_i(x'))^t} \varphi_i(x', y)\mathrm{d}y \qquad \text{for a.e. } x' \in \mathbb{R}^{n-1}, \\ i = 1, \dots, k. \tag{5.7}$$

We fix $x' \in \mathbb{R}^{n-1}$ and $i \in \{1, \dots, k\}$ such that both integrals in (5.7) converge. Assume first that $\varphi(y) := \varphi_i(x', y)$ is a step function of the form

$$\varphi = \delta\left(-C + \sum_{j=1}^{l} \chi(N_j)\right), \tag{5.8}$$

where $N_j \in \mathcal{M}(\mathbb{R})$, $N_j = N_j^*$, $j = 1, \dots, l$, $N_1 \supset \cdots \supset N_l$, $C \geq 0$, $\delta > 0$.

By the monotonicity (2.13) we have that

$$\begin{aligned} \int_{M_i(x')} \varphi(y)\mathrm{d}y &= \delta\left(-C|M_i(x')| + \sum_{j=1}^{l} |M_i(x') \cap N_j|\right) \\ &\leq \delta\left(-C|(M_i(x'))^t| + \sum_{j=1}^{l} |(M_i(x'))^t \cap N_j|\right) \\ &= \int_{(M_i(x'))^t} \varphi(y)\mathrm{d}y. \end{aligned}$$

A general φ can be approximated in $L^p(\mathbb{R})$ by step functions of the form (5.8). This proves (5.7) and thus (5.6) is established. Now in view of (5.5) we obtain (5.4).

(2) Next let $u \in L^p(\Omega)$. In view of (5.2) both integrals in (5.4) converge. We can choose an increasing sequence of step functions $\{u_m\}$ of the form (4.2) which converges to u in $L^p(\Omega)$. Then we have by (5.2)

$$\begin{aligned} &|F(x, u_m(x))| \leq A(x)B(x', u_m(x)) \leq A(x)B(x', u(x)) =: f(x), \\ &|F(x, u_m^t(x))| \leq A(x)B(x', u_m^t(x)) \leq A(x)B(x', u^t(x)) =: g(x) \qquad \forall x \in \Omega \\ &\qquad \text{and} \quad f, g \in L^1(\Omega). \end{aligned} \tag{5.9}$$

By Lemma 4.1 we can choose a subsequence $\{u_{m'}\}$ such that

$$\begin{matrix} u_{m'}(x) \longrightarrow u(x) \\ u_{m'}^t(x) \longrightarrow u^t(x) \end{matrix} \qquad \text{for a.e. } x \in \Omega. \tag{5.10}$$

In view of (5.9) and (5.10) and since $F(x, v)$ is continuous in v, (5.4) follows by Lebesgue's convergence theorem. ■

Theorem 5.2. *Let Ω be an open set with $\Omega = \Omega^*$. Let $u \in L_+^p(\mathbb{R}^n)$ for some $p \in [1, +\infty)$ and suppose that u vanishes outside Ω and satisfies* (4.20). *Furthermore, suppose that*

$f = f(x, v)$ is measurable on $\Omega \times [0, \sup u]$, symmetrically nonincreasing in y and $(p^{-1} + q^{-1} = 1)$

$$|f(x, v)| \leq a(x)b(x', v) \qquad \forall (x, v) \in \Omega \times [0, \sup u], \tag{5.11}$$

where $b(x', v)$ is nonnegative, measurable in x' and nondecreasing and right-continuous in v, and

$$a \in L_+^{q/(1-\beta)}(\Omega), b(\cdot, u(\cdot)) \in L^{q/\beta}(\Omega)$$

for some $\beta \in [0, 1]$. Finally, assume that u,f satisfy one of the conditions (i)–(iv):

(i) *$f(x, v)$ is nonincreasing in v;*
(ii) *$f(x, v)$ satisfies a Hölder condition in v with exponent $\lambda \in [1 - p^{-1}, 1]$, uniformly for every $x \in \Omega$, and $u \in L^{\lambda q}(\Omega)$;*
(iii) *$f(x, v)$ is continuous in v and bounded;*
(iv) *$f(x, v) = h(x)k(x', v)$, where $h \in L_+^{q/(1-\beta)}(\Omega)$, h is symmetrically nonincreasing in y, $k(x', v)$ is nonnegative, measurable in x' and nondecreasing in v and $k(\cdot, u(\cdot)) \in L^{q/\beta}(\Omega)$.*

Then (5.1) *holds, and in case* (i) *we have*

$$\int_\Omega f(x, u)(u^t - u)\mathrm{d}x \geq 0 \qquad \forall t \in [0, +\infty]. \tag{5.12}$$

Proof. We set

$$F(x, v) := \int_0^v f(x, w)\mathrm{d}w \qquad \forall (x, v) \in \Omega \times [0, \sup u]. \tag{5.13}$$

Since

$$|F(x, v)| \leq a(x) \int_0^v b(x', w)\mathrm{d}w \leq a(x)vb(x', v) \qquad \forall (x, v) \in \Omega \times [0, \sup u],$$

we see that F satisfies the assumptions of Theorem 5.1. Thus we derive

$$0 \leq \int_\Omega (F(x, u^t) - F(x, u))\mathrm{d}x = \int_0^1 \int_\Omega f(x, u + \theta(u^t - u))(u^t - u)\mathrm{d}x\,\mathrm{d}\theta.$$

This means that

$$\begin{aligned}\int_\Omega f(x, u)(u^t - u)\mathrm{d}x &\geq \int_0^1 \int_\Omega (f(x, u) - f(x, u + \theta(u^t - u)))(u^t - u)\mathrm{d}x\,\mathrm{d}\theta \\ &=: I(t).\end{aligned} \tag{5.14}$$

Note that in view of the assumptions on f the integral $I(t)$ converges. Now in case (i) we immediately derive (5.12) from (5.14).

Furthermore, we obtain by Hölder's inequality and by (4.20)

$$\begin{aligned}|I(t)| &\leq \|u^t - u\|_p \int_0^1 \|f(\cdot, u + \theta(u^t - u)) - f(\cdot, u)\|_q \mathrm{d}\theta \\ &\leq Ct \int_0^1 \|f(\cdot, u + \theta(u^t - u)) - f(\cdot, u)\|_q \mathrm{d}\theta.\end{aligned} \tag{5.15}$$

In view of (5.14) and (5.15) it suffices to prove that

$$\int_0^1 \|f(\cdot, u+\theta(u^t-u)) - f(\cdot, u)\|_q \mathrm{d}\theta \longrightarrow 0 \qquad \text{as } t \to 0. \qquad (5.16)$$

In the case (ii) we obtain

$$\|f(\cdot, u+\theta(u^t-u)) - f(\cdot, u)\|_q^q \le C\|u^t-u\|_{\lambda q}^{\lambda q},$$

and (5.16) follows.

Next consider case (iii), and assume that (5.16) is not true. Then there is a sequenc $t_m \searrow 0$ such that

$$\int_0^1 \|f(\cdot, u+\theta(u^{t_m}-u)) - f(\cdot, u)\|_q \mathrm{d}\theta \ge \delta \qquad (5.17$$

for some $\delta > 0$. By passing to a subsequence $\{t_{m'}\}$ we can achieve that $u^{t_{m'}}(x) \longrightarrow u(x$ a.e. in Ω as $m' \to +\infty$. This yields

$$\begin{aligned} f(x, u(x) + \theta(u^{t_{m'}}(x) - u(x))) \longrightarrow f(x, u(x)) \qquad & \text{a.e. in } \Omega, \\ \text{as } m' \to +\infty, \qquad & \theta \in [0,1]. \end{aligned} \qquad (5.18$$

If Ω is bounded, then by Lebesgue's convergence theorem (5.17) immediately yields contradiction. If Ω is unbounded, then we derive by Lebesgue's theorem

$$\int_0^1 \|f(\cdot, u+\theta(u^{t_{m'}}-u)) - f(\cdot, u)\|_{q,[\Omega\cap B_R]} \mathrm{d}\theta \longrightarrow 0 \quad \text{as } m' \to \infty \quad \forall R > 0. \qquad (5.19$$

Furthermore, from the Hardy–Littlewood inequality (2.36) we have for every $R > 0$

$$\begin{aligned} \|b(\cdot, u(\cdot))\|_{q/\beta,[\Omega\cap B_R]}^{q/\beta} &= \int_\Omega b(x', u)^{q/\beta}\chi_{B_R}\mathrm{d}x \le \int_\Omega b(x', u^t)^{q/\beta}\chi_{B_R}\mathrm{d}x \\ &= \|b(\cdot, u^t(\cdot))\|_{q/\beta,[\Omega\cap B_R]}^{q/\beta}, \qquad t \in [0, +\infty]. \end{aligned}$$

Together with the assumptions (5.11) this yields

$$\begin{aligned} &\int_0^1 \|f(\cdot, u+\theta(u^t-u)) - f(\cdot, u)\|_{q,[\Omega\setminus B_R]} \mathrm{d}\theta \\ &\le \|a\|_{q/(1-\beta),[\Omega\setminus B_R]}(\|b(\cdot, u^{t_{m'}}(\cdot))\|_{q/\beta,[\Omega\setminus B_R]} + \|b(\cdot, u(\cdot))\|_{p,[\Omega\setminus B_R]}) \\ &\le 2\|a\|_{q/(1-\beta),[\Omega\setminus B_R]}\|b(\cdot, u(\cdot))\|_{q/\beta,[\Omega\setminus B_R]} \\ &\longrightarrow 0 \quad \text{as } R \to \infty, \text{ uniformly } \forall m' \in \mathbb{N}. \end{aligned} \qquad (5.20$$

Now (5.20) together with (5.19) contradict to (5.16). Finally, in the case (iv) we hav $k(\cdot, u(\cdot)) \in L^{q/\beta}(\Omega)$ and from Definition 2.4 we infer

$$k(\cdot, u^t(\cdot)) = (k(\cdot, u(\cdot)))^t.$$

By using Lemma 4.1 this yields

$$\int_0^1 \|f(\cdot, u+\theta(u^t-u)) - f(u)\|_q \mathrm{d}\theta$$

$$\le \|h\|_{q/(1-\beta)} \int_0^1 \|k(\cdot, u+\theta(u^t-u)) - k(\cdot,u)\|_{q/\beta} \mathrm{d}\theta$$
$$\le \|h\|_{q/(1-\beta)} \|(k(\cdot,u))^t - k(\cdot,u)\|_{q/\beta} \longrightarrow 0 \qquad \text{as } t \to 0.$$

The Theorem is proved. ■

Remark 5.2. (1) The conditions (5.2) and (5.11) ensure in essence the applicability of Lebesgue's convergence theorem in the proofs, and we may replace these conditions by similar other ones. (2) Note that, if u is bounded in Theorem 5.2, then (ii) is a special case of (iii) by (5.11). Thus the case (ii) is meaningful only for unbounded functions u.

The proof of Theorem 5.2 is based on an estimate for the function u of the form (4.20). If $\Omega = \mathbb{R}^n$, then such an estimate can be ensured if $u \in W_+^{1,p}(\mathbb{R}^n)$ and if u has some decaying properties near infinity (see Theorem 4.4 and Remark 4.2). On the other hand, these estimates could be rather restrictive for some applications. Fortunately under an additional (weak) assumption on $f(x, v)$ we can bypass any strong decaying requirement for u.

Lemma 5.1. *Let $u \in W_+^{1,p}(\mathbb{R}^n) \cap C(\mathbb{R}^n)$ for some $p \in [1, +\infty]$, let $u > 0$ in $\mathbb{R}^n$ and*

$$u(x) \longrightarrow 0 \qquad \text{as } |x| \to +\infty. \tag{5.21}$$

Furthermore, let f together with u satisfy the assumptions of Theorem 5.2 with Ω replaced by $\mathbb{R}^n$ and, in addition, suppose that for some numbers $R, \delta > 0$

$$f(x,v) \quad \textit{is nonincreasing in } v \textit{ for } 0 < v < \delta \textit{ and } |x| > R. \tag{5.22}$$

Then the conclusions of Theorem 5.2 hold.

Proof. We proceed similarly as in the previous proof. First we obtain (5.14) with $\Omega = \mathbb{R}^n$. If $f(x, v)$ is nonincreasing in v, then we can argue exactly as before. In the remaining cases (ii)–(iv) we choose R_0 large enough and $R_0 > R$ such that $u \le \delta$ in $\mathbb{R}^n \setminus B_{R_0}$. Then it follows by monotonicity that $u^t \le \delta$ in $\mathbb{R}^n \setminus B_{R_0}$ for every $t \in [0, +\infty]$. Thus we have

$$I(t) \ge \int_0^1 \int_{B_{R_0}} (f(x,u) - f(x, u+\theta(u^t-u)))(u^t-u)\mathrm{d}x\,\mathrm{d}\theta =: I_1(t). \tag{5.23}$$

We choose $\varepsilon > 0$ small enough such that $u \ge \varepsilon$ in B_{R_0}. Again we have $u^t \ge \varepsilon$ in B_{R_0} for every $t \in [0, +\infty]$ by monotonicity. By setting $u_\varepsilon := \max\{u; \varepsilon\}$, we see that

$$u = u_\varepsilon,\ u^t = (u_\varepsilon)^t \quad \text{in } B_{R_0} \quad \forall t \in [0, +\infty]. \tag{5.24}$$

In view of (5.21) $(u_\varepsilon - \varepsilon)$ has bounded support, i.e. $(u_\varepsilon - \varepsilon) \in W_{0+}^{1,p}(B_{R_1})$ for some $R_1 > R_0$. An application of Theorem 4.2 to u_ε yields

$$\|(u_\varepsilon)^t - u_\varepsilon\|_p \le tR_1 \|(u_\varepsilon)_y\|_p \le tR_1 \|u_y\|_p \qquad \forall t \in [0, +\infty]. \tag{5.25}$$

Now by using (5.23)–(5.25) we compute finally

$$|I_1(t)| \le \|u^t - u\|_{p,B_{R_0}} \int_0^1 \|f(\cdot, u+\theta(u^t-u)) - f(\cdot,u)\|_{q,B_{R_0}} \mathrm{d}\theta$$
$$= \|(u_\varepsilon)^t - u_\varepsilon\|_{p,B_{R_0}} \int_0^1 \|f(\cdot, u_\varepsilon + \theta((u_\varepsilon)^t - u_\varepsilon)) - f(\cdot, u_\varepsilon)\|_{q,B_{R_0}} \mathrm{d}\theta$$

$$\leq \|(u_\varepsilon)^t - u_\varepsilon\|_p \int_0^1 \|f(\cdot, u_\varepsilon + \theta((u_\varepsilon)^t - u_\varepsilon)) - f(\cdot, u_\varepsilon)\|_q \mathrm{d}\theta$$
$$\leq tR_1 \|u_y\|_p \int_0^1 \|f(\cdot, u_\varepsilon + \theta((u_\varepsilon)^t - u_\varepsilon)) - f(\cdot, u_\varepsilon)\|_q \mathrm{d}\theta,$$

and the assertions follow by proceeding as in the previous proof.

For some applications it will be useful to have an estimate like (5.1) with the functio replaced by *any* element of the (set-valued) *maximal monotone graph* $\tilde{f}(x, v)$ of f w respect to v (compare Remark 7.1(4)). $\tilde{f}$ is defined by

$$\tilde{f}(x, v) := \left[\liminf_{h\to 0} f(x, v+h), \limsup_{h\to 0} f(x, v+h)\right] \quad \forall (x, v) \in \Omega \times [0, \sup] \qquad (5.$$

Note that, if F is defined by (5.13), then we can write alternatively

$$\tilde{f}(x, v) = \partial_v F(x, v),$$

where $\partial_v F$ is the set-valued differential of $F(x, v)$ with respect to v,

$$\partial_v F(x, v) := \left[\liminf_{h\to 0} \frac{F(x, v+h) - F(x, v)}{h}, \ \limsup_{h\to 0} \frac{F(x, v+h) - F(x, v)}{h}\right.$$
$$\forall (x, v) \in \Omega \times [0, \sup u]. \qquad (5.$$

COROLLARY 5.1

The conclusions of Theorem 5.2 *and Lemma* 5.1 *hold if the function* $f(\cdot, u(\cdot))$ *in* (5.1 *replaced by any function* g *with* $g(\cdot) \in \tilde{f}(\cdot, u(\cdot))$.

Proof. From (5.11) we obtain that $|g(x)| \leq a(x)b(x', u(x))$ $\forall x \in \Omega$, which means t $g \in L^q(\Omega)$. This ensures the convergence of the integral $\int_\Omega g(x)(u^t - u)\mathrm{d}x$. Obviou there is nothing to prove if $f(x, v)$ is continuous in v. In the remaining cases we proc as in the proof of Theorem 5.2 to infer that

$$\int_\Omega g(x)(u^t - u)\mathrm{d}x \geq \int_0^1 \int_\Omega (g(x) - f(x, u + \theta(u^t - u)))(u^t - u)\mathrm{d}x\,\mathrm{d}\theta$$
$$=: I_2(t). \qquad (5.$$

If $f(x, v)$ is nonincreasing v, then we have

$$(g(x) - f(x, u(x) + \theta(u^t(x) - u(x))))(u^t(x) - u(x)) \geq 0 \quad \forall x \in \Omega, \quad \theta \in [0$$

Together with (5.28) this leads to

$$\int_\Omega g(x)(u^t - u)\mathrm{d}x \geq 0 \qquad \forall t \in [0, +\infty]. \qquad (5.$$

Furthermore, if $f(x, z)$ is nondecreasing in z, then

$$|g(x) - f(x, u(x) + \theta(u^t(x) - u(x)))||u^t(x) - u(x)|$$
$$\leq |f(x, u^t(x)) - f(x, u(x))||u^t(x) - u(x)| \qquad \forall x \in \Omega, \quad \theta \in [0, 1].$$

In view of (5.28) we infer that

$$|I_2(t)| \leq \|u^t - u\|_p \|f(\cdot, u^t) - f(\cdot, u)\|_q.$$

Then the assertion follows by proceeding as in the proof of Theorem 5.2. ■

Remark 5.3. The reader verifies easily that Theorem 5.2, Lemma 5.1 and Corollary 5.1 hold true if the function f can be decomposed into a finite sum $\sum_{i=1}^k f_i$ where each of the functions f_i, $i = 1, \dots, k$, satisfies at least one of the conditions (i)–(iv) of Theorem 5.2.

6. Local symmetry

In this section we study functions satisfying the 'local' symmetry property (LS) from the Introduction and the relation to continuous symmetrization.

DEFINITION 6.1 (Local symmetry)

Let $u \in \mathcal{S}_+(\mathbb{R}^n)$ and continuously differentiable on $\{x : 0 < u(x) < \sup u\}$, and suppose that this last set is open. Further, suppose that u has the following property. If $x^1 = (x_0', y_1) \in \mathbb{R}^n$ with

$$0 < u(x^1) < \sup u, \quad \frac{\partial u}{\partial y}(x^1) > 0, \tag{6.1}$$

and x^2 is the (unique!) point satisfying

$$x^2 = (x_0', y_2), \quad y_2 > y_1, \quad u(x_1) = u(x^2) < u(x_0', y) \quad \forall y \in (y_1, y_2), \tag{6.2}$$

then

$$\frac{\partial u}{\partial x_i}(x^1) = \frac{\partial u}{\partial x_i}(x^2), \quad i = 1, \dots, n-1, \quad \text{and}$$
$$\frac{\partial u}{\partial y}(x^1) = -\frac{\partial u}{\partial y}(x^2). \tag{6.3}$$

Then u is called locally symmetric in the direction y.

Remark 6.1. *Geometrical meaning of local symmetry*: (1) The condition $u \in \mathcal{S}_+(\mathbb{R}^n) \cap C^1(\{0 < u < \sup u\})$ is satisfied in the following typical cases (a) and (b).

(a) $u \in C_+^1(\mathbb{R}^n)$ and $\lim_{|x| \to \infty} u(x) = 0$.
(b) There are two bounded open sets $\Omega_i \subset \mathbb{R}^n$, $i = 0, 1$, with $\Omega_0 \subset\subset \Omega_1$, $u \in C(\mathbb{R}^n \setminus \overline{\Omega_0}) \cap C^1(\Omega_1 \setminus \overline{\Omega_0})$, $0 < u < \sup u$ in $\Omega_1 \setminus \overline{\Omega_0}$, $u \equiv 0$ in $\mathbb{R}^n \setminus \overline{\Omega_1}$, $u \equiv \sup u$ in Ω_0 and

$$u(x) \longrightarrow \sup u \quad \text{if } x \to \partial\Omega_0, \quad x \in \Omega_1 \setminus \overline{\Omega_0}.$$

Note that we did not exclude $\sup u = +\infty$ in case (b). (2) Let u, x^1, x^2 be as in Definition 6.1 and let U_1 be the maximal connected component of $\{0 < u < \sup u\} \cap \{u_y > 0\}$ containing x^1. Since $u \in C^1(\{0 < u < \sup u\})$ we have that for every $(x', y) \in U_1$

$$u(x', y) = u(x', y_1 + y_2 - y) < u(x', z) \qquad \forall z \in (y, y_1 + y_2 - y). \tag{6.4}$$

The condition (6.4) says that U_1 finds a congruent counterpart after reflection about some hyperplane $\{y = \text{const}\}$. Repeating this consideration for arbitrary components of

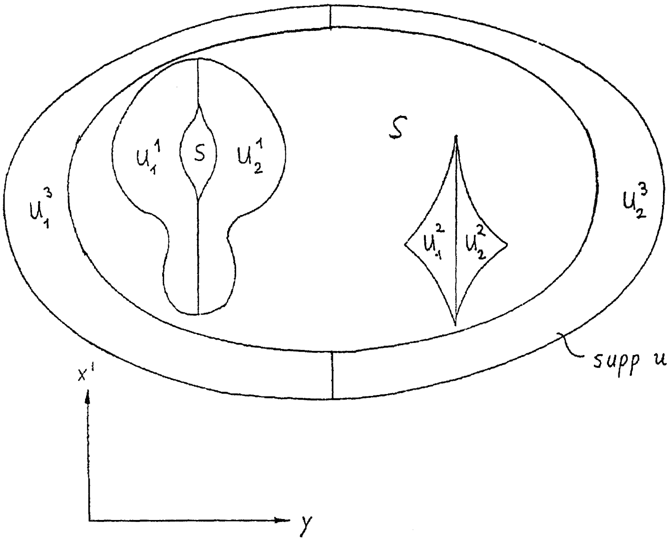


Figure 4.

$\{0 < u < \sup u\} \cap \{u_y > 0\}$ we infer the decomposition

$$\{0 < u < \sup u\} = \bigcup_{k=1}^{m} (U_1^k \cup U_2^k) \bigcup S. \qquad (6.$$

Here U_1^k is some maximal connected component of $\{0 < u < \sup u\} \cap \{u_y > 0\}$, U_2^k is reflection about some hyperplane $\{y = d_k\}$, $d_k \in \mathbb{R}$, and we have

$$u_y = 0 \qquad \text{in } S, \qquad (6.$$

and for every $(x', y) \in U_1^k$,

$$u(x', y) = u(x', 2d_k - y) < u(x', z) \quad \forall z \in (y, 2d_k - y), \quad k = 1, \ldots, m. \quad (6.$$

Note that all the sets on the right-hand side of (6.5) are disjoint and there can be countable number of U_1^k's, i.e. $m = +\infty$.

In many applications we need an 'isotropic' variant of local symmetry.

DEFINITION 6.2

Let u be as in Definition 6.1. u is called locally symmetric in every direction if for eve rotation of the cartesian coordinate system $x \longmapsto \xi = (\xi', \eta)$, $\xi' \in \mathbb{R}^{n-1}$, $\eta \in \mathbb{R}$, t function $v(\xi) := u(x)$ is locally symmetric with respect to η.

Surprisingly it can be proved that functions which are locally symmetric in *eve* direction are 'locally' *radially symmetric* (see figure 5).

Theorem 6.1. *Let u be locally symmetric in every direction. Then we have the followi decomposition,*

$$\{0 < u < \sup u\} = \bigcup_{k=1}^{m} A_k \bigcup S, \qquad (6$$

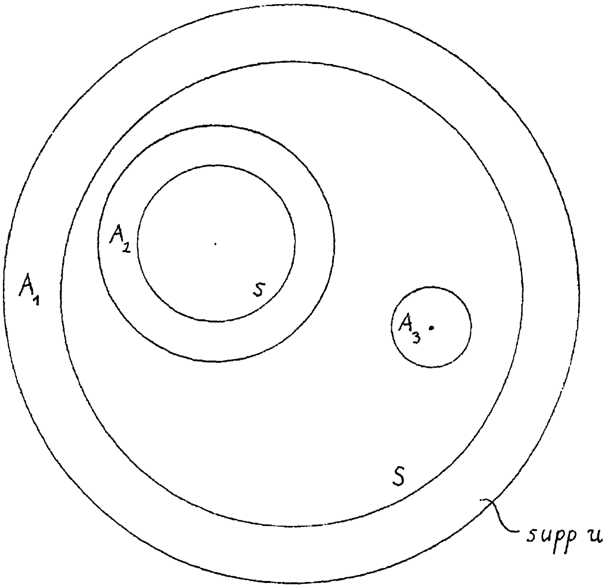


Figure 5.

where the A_k's are pairwise disjoint annuli $B_{R_k}(z_k) \setminus \overline{B_{r_k}(z_k)}$ with $R_k > r_k \geq 0$, $z_k \in \{0 < u < \sup u\}$, u is radially symmetric in A_k, and more precisely,

$$u = u(|x - z_k|), \frac{\partial u}{\partial \rho} < 0 \quad \textit{in } A_k, \tag{6.9}$$

$$(\rho = |x - z_k|), \quad \textit{and}$$

$$u(x) \geq u|_{\partial B_{r_k}(z_k)} \quad \forall x \in B_{r_k}(z_k),$$

$$1 \leq k \leq m. \tag{6.10}$$

Furthermore, we have

$$\nabla u = 0 \qquad \textit{in } S, \tag{6.11}$$

and there can be a countable number of annuli, i.e. $m = +\infty$. Finally, if $\{u > 0\}$ is unbounded, then the case $R_1 = +\infty$ is possible.

Proof. We use the notations of Definition 6.1. Let x^1, x^2 be two points which satisfy (6.1), (6.2), let U_1 be the connected component of $\{0 < u < \sup u\} \cap \{x : u_y(x) > 0\}$ containing x^1 and suppose that $\bar{x}$ is some point in U_1 with $u(x^1) = u(\bar{x})$. By a suitable rotation of the coordinate system $x \longmapsto \xi = (\xi', \eta)$, $\xi' \in \mathbb{R}^{n-1}, \eta \in \mathbb{R}$, about the point x^2 we can achieve that the ray connecting $\bar{x}$ and x^2 points into the positive η-direction, i.e. $\bar{x} \longmapsto \xi^1 = (\xi'_0, \eta_1)$ and $x^2 \longmapsto \xi^2 = (\xi'_0, \eta_2)$. We set $v(\xi) := u(x)$. It is easy to see that, if the distance $|x^1 - \bar{x}|$ is small enough (say $|x^1 - \bar{x}| < \varepsilon$), then $v_\eta(\xi^1) > 0$ and $v(\xi^1) = v(\xi^2) < v(\xi'_0, \eta)$ $\forall \eta \in (\eta_1, \eta_2)$. By the assumptions this means that

$$\frac{\partial v}{\partial \eta}(\xi^1) = -\frac{\partial v}{\partial \eta}(\xi^2) > 0 \qquad \text{and}$$

$$\frac{\partial v}{\partial \xi_i}(\xi^1) = \frac{\partial v}{\partial \xi_i}(\xi^2), \quad i = 1, \ldots, n-1.$$

From a simple computation follows that the set

$$\Gamma_1 := U_1 \cap \{x : |x - x^1| < \varepsilon \quad \text{and } u(x) = u(x^1)\}$$

is an open subset of some sphere $\{x : |x - z| = \rho_1\}$, $z \in \Omega$, $\rho_1 > 0$, and

$$\frac{\partial u}{\partial \rho}(x) = \frac{\partial u}{\partial \rho}(x^1) < 0 \quad \forall x \in \Gamma_1, \quad (\rho : \text{radial distance from } z).$$

Let $\hat{\Gamma}_1$ denote the maximal connected component of the set

$$\left\{x : \ |x - z| = \rho_1 \text{ and } \frac{\partial u}{\partial \rho}(x) < 0\right\},$$

containing the point x^1. Then, proceeding as before, we obtain that $\hat{\Gamma}_1$ is relatively open in $\{x : |x - z| = \rho_1\}$ and

$$\frac{\partial u}{\partial \rho}(x) = \frac{\partial u}{\partial \rho}(x^1) \qquad \forall x \in \hat{\Gamma}_1.$$

This means that $\hat{\Gamma}_1$ is relatively closed in $\{x : |x - z| = \rho_1\}$. Thus we have

$$\hat{\Gamma}_1 = \{x : |x - z| = \rho_1\}.$$

We can repeat these arguments for all points of U_1. Since $u \in C^1(\{0 < u < \sup u\})$ we infer that u is radially symmetric in $B_R(z) \setminus \overline{B_r(z)}$ for some $R > r \geq 0$ and $(\partial u/\partial \rho)(x) < 0 \ \forall x \in B_R(z) \setminus \overline{B_r(z)}$. Note that, if $\{u > 0\}$ is unbounded, then possibly $R = +\infty$. The Theorem is proved. ■

Next we give a purely *analytic* description of local symmetry in terms of continuous symmetrization.

Theorem 6.2. *Let Ω be an open set with $\Omega = \Omega^*$ and $u \in W_{0+}^{1,p}(\Omega)$ for some $p \in [1, +\infty)$. Further, let G be a strictly convex Young function satisfying* (3.43). *Finally, let u be continuously differentiable on $\{x : 0 < u(x) < \sup u\}$ and suppose that this last set is open. Then, if*

$$\lim_{t \searrow 0} \frac{1}{t}\left(\int_{\mathbb{R}^n} G(|\nabla u|)\mathrm{d}x - \int_{\mathbb{R}^n} G(|\nabla u^t|)\mathrm{d}x\right) = 0, \tag{6.12}$$

u is locally symmetric in direction y.

Proof. Let x^1, x^2 satisfy (6.1) and (6.2) and let U_1 be as in Remark 6.1. We have $u_y(x^2) \leq 0$. First assume that $u_y(x^2) < 0$. There are small neighbourhoods W_1, W_2 of the points x^1 and x^2, respectively, such that $u_y > 0$ in W_1 and $u_y < 0$ in W_2. Let $y_i = y_i(x', u)$, $i = 1, 2$, denote the corresponding inverse functions which exist for every (x', u) lying in a small neighbourhood V of the point $(x'_0, u(x'_0, y_1))$. Then the function u^t can be represented by corresponding inverse functions y_i^t, $i = 1, 2$, according to the formulas (2.10) for sufficiently small values $t > 0$, say $0 < t < t_0$. Let $G_i(t)$ denote the images of V in the (x', y)-domain after the mappings $(x', u) \longmapsto (x', y_i^t(x', u))$, $i = 1, 2$. Note that $G_1(0) \subset W_1$ and $G_2(0) \subset W_2$.

We approximate u by good functions which coincide with u in the domains $G_i(0)$, $i = 1, 2$. By proceeding as in the proof of Theorem 3.1 we infer that

$$\int_{\Omega\setminus(G_1(0)\cup G_2(0))} G(|\nabla u|)\mathrm{d}x \geq \int_{\Omega\setminus(G_1(t)\cup G_2(t))} G(|\nabla u^t|)\mathrm{d}x \tag{6.13}$$

and

$$\begin{aligned} I(t) &:= \int_{G_1(t)\cup G_2(t)} G(|\nabla u^t|)\mathrm{d}x \\ &= \sum_{k=1}^{2} \int_V G\left(\left\{1 + \sum_{i=1}^{n-1} \left(\frac{\partial y_k^t}{\partial x_i}\right)^2\right\}^{1/2} \left|\frac{\partial y_k^t}{\partial u}\right|^{-1}\right) \left|\frac{\partial y_k^t}{\partial u}\right| \mathrm{d}x'\mathrm{d}u \\ &\quad \forall t \in (0, t_0). \end{aligned} \tag{6.14}$$

We introduce the parameter $\lambda := (1/2)(1 - \mathrm{e}^{-t})$, $t \in [0, +\infty]$, and set $\psi(\lambda) := I(t)$. By setting $\psi(1 - \lambda) := \psi(\lambda)\ \forall\lambda \in [0, (1/2)]$, we formally extend the definition of $\psi(\lambda)$ for all $\lambda \in [0, 1]$. Assume for a moment that $\psi(0) > \psi(1/2)$. Since $\psi(\lambda)$ is convex we obtain that $\lim_{t\searrow 0}(I(t) - I(0))/t < 0$. In view of (6.13) and (6.14) this contradicts to (6.12). Thus we have $\psi(0) = \psi(1/2)$. Since G is *strictly* convex we infer from this that $y_{1,x_i} = y_{2,x_i}$, $i = 1, \dots, n-1$, and $y_{1,u} = -y_{2,u}$ almost everywhere in V. This means that (6.4) is satisfied throughout the domain $G_1(0)$.

Next let us assume that $u_y(x^2) = 0$. Since $u_y(x^1) > 0$, the implicit function theorem tells us that the problem

$$u(x_0', y) = u(x_0', y_1) + \varepsilon, \quad (x_0', y) \in G_1(0),$$

has a unique solution $y = y_1^\varepsilon$ if ε is positive and small enough, say $\varepsilon \in (0, \varepsilon_0)$. For $\varepsilon \in (0, \varepsilon_0)$ let y_2^ε denote the (unique!) number satisfying $y_1^\varepsilon < y_2^\varepsilon$ and $u(x_0', y_1^\varepsilon) = u(x_0', y_2^\varepsilon) < u(x_0', y)\ \forall y \in (y_1^\varepsilon, y_2^\varepsilon)$. Since u is differentiable we can choose a sequence $\varepsilon_m \searrow 0$ such that $u_y(x_0', y_2^{\varepsilon_m}) < 0$. Then from the earlier considerations follows that $u_y(x_0', y_1^{\varepsilon_m}) = -u_y(x_0', y_2^{\varepsilon_m})$. Clearly we have $\lim_{m\to\infty} y_i^{\varepsilon_m} = y_i$, $i = 1, 2$. Since $u \in C^1(\{0 < u < \sup u\})$ this yields

$$\lim_{m\to\infty} u_y(x_0', y_2^{\varepsilon_m}) = -u_y(x_0', y_1) < 0,$$

a contradiction. Thus the condition (6.4) is again satisfied for all $x = (x', y) \in G_1(0)$. Now set

$$\hat{G}_1 := \{(x', y) \in U_1\ :\ u(x', y) = u(x', y_1 + y_2 - y) < u(x', z)\ \forall z \in (y, y_1 + y_2 - y)\}.$$

Obviously we have $G_1(0) \subseteq \hat{G}_1$, and we can argue as before to infer that $\hat{G}_1$ is relatively open in U_1. Let $\overline{x_m} = (\overline{x_m'}, \overline{y_m})$, $m = 1, 2, \dots$, be any sequence in $\hat{G}_1$ converging to some point $\overline{x} = (\overline{x'}, \overline{y}) \in U_1$. Since $u \in C^1(\{0 < u < \sup u\})$ we have $u_y(\overline{x}) > 0$ and $u(\overline{x}) = u(\overline{x'}, y_1 + y_2 - \overline{y}) \leq u(\overline{x'}, y)\ \ \forall y \in (\overline{y}, y_1 + y_2 - \overline{y})$. Therefore we find some value $\hat{y} \in (\overline{y}, y_1 + y_2 - \overline{y}]$ such that $u(\overline{x}) = u(\overline{x'}, \hat{y}) < u(\overline{x'}, y)\ \forall y \in (\overline{y}, \hat{y})$. If $\hat{y} < y_1 + y_2 - \overline{y}$, then $u_y(\overline{x'}, \hat{y}) = 0$. This is impossible by the earlier considerations. Thus $\hat{y} = y_1 + y_2 - y$, i.e. $\overline{x} \in \hat{G}_1$. Therefore $\hat{G}_1$ is relatively closed with respect to U_1. But this means that $\hat{G}_1 = U_1$. The Theorem is proved. ■

Analogously one can prove the following extension of Theorem 6.2.

COROLLARY 6.1

Let u satisfy the assumptions of Theorem 6.2 *with the corresponding integrals replaced by the more general ones in* (3.18), *where the functions G, a, a_{ij}, $i, j = 1, \ldots, n-1$, are as in Corollary* 3.3 *and $G(x', v, z)$ is strictly convex in z. Then the conclusions of Theorem* 6.2 *hold.*

7. Elliptic problems

Now we apply the preceding considerations to elliptic problems. First we deal with the variational problem (P) from the introduction.

Theorem 7.1. *Let Ω be a domain in $\mathbb{R}^n$ with $\Omega = \Omega^*$. For some $p \in [1, +\infty)$ let K be a closed subset of $W^{1,p}_{0+}(\Omega)$ and assume that K has the property that, if $v \in K$, then also $v^t \in K$ for every $t \in [0, +\infty]$. Let $G = G(x', v, z)$ be nonnegative and continuous on $\mathbb{R}^{n-1} \times (\mathbb{R}^+_0)^2$, and suppose that G is strictly convex in z and satisfies* (3.34). *Furthermore, let u be a local minimizer of problem* (P), *and suppose that F, u satisfy the assumptions of Theorem* 5.1. *Finally assume that the set $\{0 < u < \sup u\}$ is open and $u \in C^1(\{0 < u < \sup u\})$.*

Then u is locally symmetric in direction y.

Proof. From Theorem 5.1 we infer (5.4). In view of Corollary 3.3 and the inequality $J(u) \leq J(u^t)$, $(t \in [0, +\infty])$, we obtain that

$$\int_\Omega G(x', u, |\nabla u|)\mathrm{d}x = \int_\Omega G(x', u^t, |\nabla u^t|)\mathrm{d}x \qquad \forall t \in [0, +\infty].$$

By Corollary 6.1 this means that u is locally symmetric in direction y. ■

By Corollary 3.3 and 6.1 the following extension of Theorem 7.1 is obvious.

COROLLARY 7.1

Let Ω, K, G, F, u be as in Theorem 7.1 *and, in addition, suppose that $|\nabla v|$ in* (1.1) *is replaced by the "generalized gradient" in* (3.18) *and the functions a, a_{ij} are as in Corollary* 3.3. *Then the conclusion of Theorem* 7.1 *holds.*

Theorem 7.1 yields the following Corollary 7.2 in the radially symmetric case.

COROLLARY 7.2

Let $\Omega = B_R$ for some $R > 0$ or $\Omega = \mathbb{R}^n$, and let K, G, F and u be as in Theorem 7.1. *In addition, suppose that G and the function B in* (5.2) *are independent of x, F satisfies* (5.3) *in every rotated coordinate system and*

$$F = F(|x|, v) \qquad \forall (x, v) \in \Omega \times [0, \sup u]. \tag{7.1}$$

Then u is locally symmetric in every direction.

Remark 7.1. (1) If F is independent of x, then we may relax the conditions on F in Theorem 7.1 and Corollaries 7.1 and 7.2. By Remark 5.2 it is enough to demand in this case that F is a Borel function and $F(u) \in L^1(\mathbb{R}^n)$. (2) If F is as in Corollary 7.2 and is differentiable in v, then $(\partial F)/(\partial v)$ is nondecreasing in $|x|$. (3) In view of (2.30), the assumption on K in Theorem 7.1 means that K may include side constraints of the types $(\lambda \in \mathbb{R}, c \geq 0, \mu > 0)$

$$\varphi \leq v \leq \psi, \qquad \text{where } \varphi = \varphi^*, \psi = \psi^*, \tag{7.2}$$

$$\int_\Omega g(v)\mathrm{d}x = \lambda, \qquad \text{where } g \text{ is a Borel function, or} \tag{7.3}$$

$$|\{v > c\}| = \mu. \tag{7.4}$$

In the case of the constraints (7.2) the statement of the problem allows to deal with 'ring-shaped' geometries. Note also that by the monotonicity of continuous symmetrization we infer from (7.2) that $\varphi = \varphi^t \leq v^t \leq \psi^t = \psi$, $(t \in [0, +\infty])$. Constraints of type (7.4) lead to variational solutions of *overdetermined* boundary value problems (see [Se] and [AC]). (4) Assume that $F(x, v)$ is Lipschitz continuous in v and

$$K = W_0^{1,p}(\Omega) \cap \{\text{constraints of the form (7.2)}\}.$$

Then K is *convex* and well-known analysis shows that a local minimizer u of (P) is a solution of the following (local) *variational inequality*

$$\int_\Omega G_z(x', u, |\nabla u|)|\nabla u|^{-1}\nabla u \nabla(v - u)\mathrm{d}x \geq \int_\Omega g(x)(v - u)\mathrm{d}x$$
$$\forall v \in K \text{ with } \|\nabla(v - u)\|_p < \varepsilon. \tag{7.5}$$

Here $g(\cdot) \in \partial_v F(\cdot, u(\cdot))$, $\partial_v F(x, v)$ is defined by (5.27) and ε is a given (small) constant. These well-known problems appear in models for reaction and diffusion processes (see [Di] and [K1]).

Remark 7.1(4) suggests to investigate directly the following *differential inclusion* instead of problem (P).

$$u \in W_{0+}^{1,p}(\Omega),$$
$$-\nabla(G_z(x, u, |\nabla u|)|\nabla u|^{-1}\nabla u) \in \tilde{f}(x, u). \tag{7.6}$$

Here $\tilde{f}(x, v)$ denotes the maximal monotone graph of $f(x, v)$ with respect to v (see (5.26)). The idea in proving symmetry results consists in using a Green-type identity with test function $(u^t - u)$ (namely (7.10) below) and to exploit the estimates of § 5 for small t. Clearly the assumptions on the data of the problem (7.6) and its solution will be more restrictive than in Theorem 7.1, especially in the case of unbounded domains.

Theorem 7.2. *Let Ω be a bounded domain in $\mathbb{R}^n$ with $\Omega = \Omega^*$, and let $G = G(x', z)$ be nonnegative, continuous in x', differentiable and strictly convex in z, and satisfies*

$$G(x', 0) = 0 \qquad \textit{and}$$
$$G_z(x', z) \leq C(1 + z^{p-1}) \qquad \forall (x', z) \in \mathbb{R}^{n-1} \times \mathbb{R}_0^+ \tag{7.7}$$

for some $p \in [1, +\infty)$ and $C > 0$. Furthermore, let $u \in W_{0+}^{1,p}(\Omega)$, let $f = f(x, v)$ measurable on $\Omega \times [0, \sup u]$, symmetrically nonincreasing in y and satisfies (5.11), *and suppose that f can be decomposed as follows*

$$\begin{aligned} &f = f_1 + f_2 + f_3, \qquad \textit{where} && (7.\\ &f_1 = f_1(x,v) \quad \textit{is continuous in } v, &&\\ &f_2 = f_2(x,v) \quad \textit{is nonincreasing in } v \textit{ and} &&\\ &f_3 = h(x)k(x',v), \quad \textit{with } h \textit{ and } k \textit{ as in Theorem } 5.2(\text{iv}). && \end{aligned}$$

Finally let $u \in W_{0+}^{1,p}(\Omega)$ *satisfy weakly*

$$-\nabla(G_z(x',|\nabla u|)|\nabla u|^{-1}\nabla u) = g \qquad \textit{in } \Omega, \qquad (7.$$

where $g(\cdot) \in \tilde{f}(\cdot,u(\cdot))$, $\tilde{f}(x,v)$ *denotes the maximal monotone graph of* $f(x,v)$ *wi respect to* v *(see* (5.26)*). In addition, suppose that the set* $\{0 < u < \sup u\}$ *is open a* $u \in C^1(\{0 < u < \sup u\})$. *Then* u *is locally symmetric in direction* y.

Proof. Let q be defined by $p^{-1} + q^{-1} = 1$. Since $u \in W_0^{1,p}(\Omega)$ we have $f(\cdot,u(\cdot)) \in L^q($ by our assumptions and thus also $g \in L^q(\Omega)$. From (7.10) we obtain the identity

$$\int_\Omega G_z(x',|\nabla u|)|\nabla u|^{-1}\nabla u \nabla(u^t - u)\mathrm{d}x = \int_\Omega g(x)(u^t - u)\mathrm{d}x \quad \forall t \in [0,+\infty].$$

(7.1

By using the convexity of G with respect to z and Corollary 3.3, we infer from this

$$0 \geq \int_\Omega (G(x',|\nabla u^t|) - G(x',|\nabla u|))\mathrm{d}x \geq \int_\Omega g(x)(u^t - u)\mathrm{d}x \quad \forall t \in [0,+\infty].$$

(7.1

We can estimate the right-hand side of (7.11) according to Theorem 5.2. This leads t

$$\lim_{t \searrow 0} \frac{1}{t} \int_\Omega (G(x',|\nabla u^t|) - G(x',|\nabla u|))\mathrm{d}x = 0.$$

Then the assertion follows by applying Corollary 6.1.

A similar result holds also for solutions of (7.9) in the entire space.

Theorem 7.3. *Let* $G = G(x',z)$ *be nonnegative, continuous in* x', *differentiable a strictly convex in* z *and let* G *satisfy*

$$\begin{aligned} &G(x',0) = 0 \qquad \textit{and} &&\\ &G_z(x',z) \leq Cz^{p-1} \qquad \forall (x',z) \in \mathbb{R}^{n-1} \times \mathbb{R}_0^+ && (7.1 \end{aligned}$$

for some $p \in [1,+\infty)$ *and* $C > 0$. *Furthermore, let* f, u *be as in Theorem* 7.2 *w* $\Omega = \mathbb{R}^n$. *In addition, suppose that one of the following conditions* (i) *or* (ii) *is satisfie*

(i) *u satisfies the decaying properties* (4.17)–(4.19) *of Theorem* 4.3;
(ii) f *satisfies* (5.22) *and* u *is positive and satisfies* (5.21).

Then u *is locally symmetric in direction* y.

Proof. By Green's formula we obtain for every $r > 0$

$$\int_{B_r} G_z(x',|\nabla u|)|\nabla u|^{-1}\nabla u \nabla(u^t - u)\mathrm{d}x$$

$$= \int_{B_r} g(x)(u^t - u)\mathrm{d}x + \int_{\partial B_r} G_z(x', |\nabla u|)|\nabla u|^{-1}\frac{\partial u}{\partial \nu}(u^t - u)\mathrm{d}S$$
$$\forall t \in [0, +\infty], \quad (\nu : \text{exterior normal}). \tag{7.13}$$

Furthermore, we have by Hölder's inequality,

$$I(r) := \Big| \int_{\partial B_r} G_z(x', |\nabla u|)|\nabla u|^{-1}\frac{\partial u}{\partial \nu}(u^t - u)\mathrm{d}S\Big| \leq C\|\nabla u\|^{p-1}_{p,\partial B_r}\|u^t - u\|_{p,\partial B_r}, \tag{7.14}$$

for some number $C > 0$. Since $u \in W^{1,p}(\mathbb{R}^n) \cap C^1(\mathbb{R}^n)$ this means that $\lim_{r\searrow 0} I(r) = 0$. Hence, by passing to the limit $r \to +\infty$ in (7.13) we see that (7.11) holds with $\Omega = \mathbb{R}^n$. Then we argue as in the previous proof by applying Corollary 3.3, Theorem 5.2 and Corollary 6.1. in the case (i) and by using Lemma 5.1 in the case (ii). ■

Our method is also applicable to problems in ring-shaped domains.

COROLLARY 7.3

Let Ω, Ω_0 be two bounded domains in $\mathbb{R}^n$ with $\Omega = \Omega^*$, $\Omega_0 = \Omega_0^*$ and $\Omega_0 \subset\subset \Omega$. Furthermore, let G, f be as in Theorem 7.2 *and let $u \in W_0^{1,p}(\Omega)$ be a weak solution of the following problem*

$$-\nabla(G_z(x', |\nabla u|)|\nabla u|^{-1}\nabla u) = g, \quad 0 \leq u \leq 1 \quad \text{in } \Omega \setminus \Omega_0,$$
$$u \equiv 1 \quad \text{in } \overline{\Omega_0}, \tag{7.15}$$

where g is as in Theorem 7.2. *In addition, suppose that the set $\{0 < u < 1\}$ is open and that $u \in C^1(\{0 < u < 1\})$. Then u is locally symmetric in direction y.*

Proof. Since $\Omega_0 = \Omega_0^*$, we see that $(u^t - u) \in W_0^{1,p}(\Omega \setminus \overline{\Omega_0})$ for every $t \in [0, +\infty]$. Therefore the identity (7.11) again holds and we can proceed exactly as in the proof of Theorem 7.2. ■

The proof of the following corollary is analogous.

COROLLARY 7.4

Let G, f, u be as in Corollary 7.3 *with Ω replaced by $\mathbb{R}^n$. In addition, suppose that one of the conditions* (i) *or* (ii) *of Theorem* 7.3 *is satisfied. Then u is locally symmetric in direction y.*

By using the Corollaries 3.3 and 6.1 we may extend the previous results to more general differential operators (with obvious changes in the proof).

COROLLARY 7.5

Let the functions G, a, a_{ij}, $i,j = 1, \ldots, n-1$, be as in Corollary 3.4 *and independent of v. Furthermore, let f, u be as in Theorem* 7.2 *or* 7.3 *with the equation* (7.9) *replaced by*

$$-\sum_{i=1}^{n}\frac{\partial}{\partial x_i}\left(G_z\left(x', \left\{\sum_{j,k=1}^{n} a_{jk}u_{x_j}u_{x_k}\right\}^{1/2}\right)\left\{\sum_{j,k=1}^{n} a_{jk}u_{x_j}u_{x_k}\right\}^{-1/2}\sum_{j=1}^{n} a_{ij}u_{x_j}\right) = g. \tag{7.16}$$

(Here $x_n := y$, $a_{nn} := a^2$ and $a_{in} = a_{ni} := 0$, $i = 1, \ldots, n-1$.) Then the conclusions of Theorem 7.2 or 7.3, respectively, hold.

In the 'isotropic' cases the following consequences of the above results are immediate.

COROLLARY 7.6

Let u satisfy the assumptions of Theorem 7.2, 7.3 or of Corollary 7.3, 7.4. Suppose that the function G and the functions b and k in (5.10), *respectively* (7.8), *are independent of x. Further, let $\Omega = B_R$ or $\Omega = \mathbb{R}^n$ and $\Omega_0 = B_r$ for some numbers $R > r > 0$, and suppose that*

$$f = f(|x|, v), \qquad f \text{ is nonincreasing in } |x|. \tag{7.17}$$

Then u is locally symmetric in every direction.

Let us give a typical example with discontinuous nonlinearity f which is covered by Theorem 7.3 and Corollary 7.6.

Example 7.1. Let $u \in W^{1,p}(\mathbb{R}^n)$ for some $p \in (1, +\infty)$, and let $\varphi = \varphi(x')$ be a measurable function on $\mathbb{R}^{n-1}$ satisfying

$$\varphi(x') \geq \delta \qquad \forall x' \in \mathbb{R}^{n-1} \quad \text{for some } \delta > 0. \tag{7.18}$$

Further let $u \in W^{1,p}(\mathbb{R}^n)$ satisfy

$$-\Delta_p u \equiv -\nabla(|\nabla u|^{p-2}\nabla u) = g, \quad u > 0 \qquad \text{in } \mathbb{R}^n, \tag{7.19}$$

where

$$g(x) \begin{cases} = 1 & \text{if } u(x) > \varphi(x') \\ \in [0,1] & \text{if } u(x) = \varphi(x') \\ = 0 & \text{if } u(x) < \varphi(x') \end{cases} \qquad \forall x \in \mathbb{R}^n. \tag{7.20}$$

From (7.20) we see that $g(\cdot) \in \tilde{f}(\cdot, u(\cdot))$, where $\tilde{f}(x, v)$ is the maximal monotone graph of

$$f(x, v) = \chi(\{v > \varphi(x')\}, \quad (x, v) \in \mathbb{R}^n \times [0, \sup u].$$

Note that, if $p = 2$, $n = 3$ and $\varphi = \varphi(|x'|)$, then the problem (7.19) can be seen as a model for an equilibrium configuration of incompressible axially symmetric rotating fluids or rotating stars. The fluid rotates about the y-axis, the function $f(\cdot, u(\cdot))$ represents the mass density of the fluid and the function φ comes from the (prescribed) rotational law (see [Lio2, F], [B3]).

In view of (7.18) and since $u \in L^p(\mathbb{R}^n)$ we have $g \in L^1(\mathbb{R}^n)$. Since g is bounded, this yields $\lim_{|x|\to\infty} u(x) = 0$. By (7.18) we infer that g has bounded support. Now we see that u satisfies the assumptions (and in particular (ii)) of Theorem 7.3. In the particular case $\varphi \equiv \delta$, (i.e. $f(x, v) = \chi(\{v > \delta\})$, u satisfies the assumptions of Corollary 7.6.

Acknowledgements

I thank C Bandle, H Berestycki, G Buttazzo, A Cianchi, L Damascelli, A Henrot, F Pacella, W Reichel and A Yu Solynin for helpful discussions and comments. In particular, I am grateful to B Kawohl for his constant interest and encouragement during the preparation of this work.

References

[AL] Almgren J A and Lieb E H, Symmetric decreasing rearrangement is sometimes continuous. *J. Am. Math. Soc.* **2** (1989) 683–773

[Alt] Alt H-W, *Lineare Funktionalanalysis.* 2nd ed. (Berlin: Springer-Verlag) (1992)

[AC] Alt H-W and Caffarelli L, Existence and regularity for a minimum problem with free boundary. *J. Reine Angew. Math.* **325** (1981) 105–144

[ALT] Alvino A, Lions P-L and Trombetti G, On optimization problems with prescribed rearrangements. *Nonlinear Anal. T.M.A.* **13** (1989) 185–220

[BaN] Badiale M and Nabana E, A note on radiality of solutions of p-Laplacian equation. *Appl. Anal.* **52** (1994) 35–43

[Ba] Baernstein II A, A unified approach to symmetrization, in: Partial differential equations of elliptic type (eds) A Alvino *et al.* (1995) *Symposia matematica* (Cambridge Univ. Press) vol. 35 pp. 47–91

[BM] Bandle C and Marcus M, Radial averaging transformations and generalized capacities. *Math. Z.* **145** (1975) 11–17

[Be] Beckner W, Sobolev inequalities, the Poisson semigroup and analysis on the sphere S^n. *Proc. Nat. Acad. Sci. U.S.A.* **89** (1992) 4816–4819

[BeN] Berestycki H and Nirenberg L, The method of moving planes and the sliding method. *Bol. Soc. Brasil. Mat. (N.S.)* **22** (1991) 1–37

[BLL] Brascamp H J, Lieb E H and Luttinger J M, A general rearrangement inequality for multiple integrals. *J. Funct. Anal.* **17** (1974) 227–237

[B1] Brock F, Axially symmetric flow with finite cavities. Z. *Anal. Anw.* **12** (1993) I: 97–112, II: 297–303

[B2] Brock F, Continuous Steiner-symmetrization. *Math. Nachr.* **172** (1995) 25–48

[B3] Brock F, Continuous polarization and Symmetry of Solutions of Variational Problems with Potentials. in: Calculus of variations, applications and computations, Pont-a-Mousson 1994 (eds) C Bandle *et al*, *Pitman Research Notes in Math.* **326** (1995) 25–35

[B4] Brock F, *Continuous rearrangement and symmetry of solutions of elliptic problems*, Habilitation Thesis (Leipzig) (1998) pp. 124

[B5] Brock F, Weighted Dirichlet-type inequalities for Steiner-symmetrization. *Calc. Var.* **8** (1999) 15–25

[B6] Brock F, Radial symmetry for nonnegative solutions of semilinear elliptic problems involving the p-Laplacian, in: Progress in partial differential equations, Pont-á-Mousson 1997, vol. 1 (eds) H Amann *et al*, *Pitman Research Notes in Math.* **383**, 46–58

[BS] Brock F and Solynin A Yu, An approach to symmetrization via polarization. preprint, (Köln) (1996) pp. 60, to appear in *Trans. A.M.S.*

[BZ] Brothers J and Ziemer W P, Minimal rearrangements of Sobolev functions. *J. Reine Angew. Math.* **384** (1988) 153–179

[Bu] Burchard A, Steiner symmetrization is continuous in $W^{1,p}$. *GAFA* **7** (1997) 823–860

[C] Coron J-M, The continuity of rearrangement in $W^{1,p}(\mathbb{R})$. *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* **11(4)** (1984) 57–85

[Dam] Damascelli L, Comparison theorems for some quasilinear degenerate elliptic operators and applications to symmetry and monotonicity results. *Ann. Inst. H. Poincaré, Anal. non lineáire* **15** (1998) 493–516

[DamPa] Damascelli L and Pacella F, Monotonicity and symmetry of solutions of p-Laplace equations, $1 < p < 2$, via the moving plane method (1998) pp. 22 to appear in: *Ann. Scuola Norm. Sup. Pisa.*

[Da] Dancer E N, Some notes on the method of moving planes. *Bull. Austr. Math. Soc.* **46** (1992) 425–434

[Di] Diaz J I, Nonlinear PDE and free boundaries, vol. I, Elliptic equations. *Pitman Research Notes* (Boston) (1985) vol. 106

[Du] Dubinin V N, Capacities and geometric transformations in n-space. *GAFA* **3** (1993) 342–369

[EG] Evans L C and Gariepy R F, Measure theory and fine properties of functions. (London: CRC Press) (1992)

[F] Friedman A, *Variational principles and free-boundary problems* (NY: Wiley-Interscience) (1982)

[GNN] Gidas B, Ni W-M and Nirenberg L, Symmetry and related properties via the maximum principle. *Comm. Math. Phys.* **68** (1979) 209–243

[GT] Gilbarg D and Trudinger N S, Elliptic Partial Differential Equations of Second Order. 2nd ed., (Berlin: Springer-Verlag) (1983)

[GKPR] Grossi M, Kesavan S, Pacella F and Ramaswamy M, *Symmetry of positive solutions of some nonlinear equations.* preprint (Roma) (1995)

[K1] Kawohl B, Rearrangements and convexity of level sets in PDE. *Springer Lecture Notes* (1985) vol. 1150

[K2] Kawohl B, On the simple shape of stable equilibria, in: Geometry of Solutions to Partial Differential Equations (ed.) G Talenti (Academic Press) (1989) 73–89

[K3] Kawohl B, On the isoperimetric nature of a rearrangement inequality and its consequences for some variational problems. *Arch. Rat. Mech. Anal.* **94** (1986) 227–243

[K4] Kawohl B, On starshaped rearrangement and applications. *Trans. Amer. Math. Soc.* **296** (1986) 377–386

[KP] Kesavan S and Pacella F, Symmetry of positive solutions of a quasilinear elliptic equation via isoperimetric inequalities. *Appl. Anal.* **54** (1994) 7–37

[Lio1] Lions P-L, Two geometrical properties of solutions of semilinear problems. *Appl. Anal.* **12** (1981) 267–272

[Lio2] Lions P-L, Minimization problems in $L^1(\mathbb{R}^3)$. *J. Funct. Anal.* **41** (1981) 236–275

[M] Marcus M, Radial averaging of domains, estimates for Dirichlet integrals and applications. *J. d' Analyse* **27** (1974) 47–93

[McN] McNabb A, Partial Steiner symmetrization and some conduction problems. *J. Math. Anal. Appl.* **17** (1967) 221–227

[Sa] Sarvas J, Symmetrization of condensers in n-space. *Ann. Acad. Sci. Fenn. Ser. A1* **522** (1972) 1–44

[Se] Serrin J, A symmetry problem in potential theory. *Arch. Rat. Mech. Anal.* **43** (1971) 304–318

[SeZ] Serrin J and Zou H, Symmetry of ground states of quasilinear elliptic equations, preprint (1998) pp. 24

[So] Solynin A Yu, Continuous symmetrization of sets. *Zapiski Nauchnykh Seminarov LOMI Akademii Nauk SSSR* **185** (1990) 125–139

Proc. Indian Acad. Sci. (Math. Sci.), Vol. 110, No. 2, May 2000, pp. 205–211.
© Printed in India

Diameter preserving linear maps and isometries, II

FÉLIX CABELLO SÁNCHEZ

Departamento de Matemáticas, Universidad de Extremadura, Avenida de Elvas 06071-Badajoz, Spain
E-mail address: fcabello@unex.es

MS received 24 November 1999; revised 1 February 2000

Abstract. We study linear bijections of simplex spaces $\mathcal{A}(S)$ which preserve the diameter of the range, that is, the seminorm $\varrho(f) = \sup\{|f(x) - f(y)| : x, y \in S\}$.

Keywords. Isometry; simplex space; linear preserver problem.

1. Introduction and statement of the results

In this paper we study diameter preserving mappings on spaces of affine functions. Precisely, let S be a compact convex set in a locally convex Hausdorff space and let $\mathcal{A}(S)$ be the space of all (real or complex) continuous affine functions on S. We are interested in linear bijections on $\mathcal{A}(S)$ which preserve the diameter of the range, that is, the seminorm

$$\varrho(f) = \sup\{|f(x) - f(y)| : x, y \in S\}.$$

Our main result reads as follows:

Theorem 1. *Let S be a simplex. A linear bijection $T : \mathcal{A}(S) \longrightarrow \mathcal{A}(S)$ is diameter preserving if and only if there is an affine automorphism $\varphi : S \longrightarrow S$, a linear functional $\mu : \mathcal{A}(S) \longrightarrow \mathbb{K}$ and a number τ with $|\tau| = 1$ and $\mu(1_S) + \tau \neq 0$ such that $Tf = \tau f \circ \varphi + \mu(f)1_S$ for every $f \in \mathcal{A}(S)$.*

Simplexes constitute the simplest class of convex sets (see [6, 1] or [9] for precise definitions). In finite dimensional spaces, simplexes are the usual objects (segments, triangles, etc.), but an infinite dimensional simplex may be a very complex object (see the monster constructed by Poulsen in [7]).

Diameter preserving bijections on spaces of continuous functions have been recently studied by a number of authors (see the papers [4, 3, 2]; for an introduction to linear preserver problems we suggest the survey paper [5]). Of course, by a diameter preserving mapping on $C(X)$ (the space of real or complex continuous functions on the compact Hausdorff space X) we mean a mapping which preserves the seminorm

$$\varrho_X(f) = \sup\{|f(x) - f(y)| : x, y \in X\}.$$

The basic result for $C(X)$ spaces is the following ([4], Theorem, [2], Theorem 1, [3], Theorem 5.1):

Theorem 2. *Let X be a compact Hausdorff space. A linear bijection T of $C(X)$ is diameter preserving if and only if there is a homeomorphism ϕ of X, a linear functional*

$\mu : C(X) \to \mathbb{K}$ and a number τ with $|\tau| = 1$ and $\mu(1_X) + \tau \neq 0$ such that $Tf = \tau f \circ \phi + \mu(f)1_X$ for every $f \in C(X)$.

We feel that Theorem 1 puts this result in its proper setting. Let us explain why. It is well-known (see [9] or [1]) that the space $C(X)$ can be regarded as the space of continuous affine functions on the set of all regular Borel probabilities on X

$$S = \{\mu \in M(X) : \mu(X) = |\mu|(X) = 1\}$$

endowed with the relative weak* topology of $M(X)$ viewed as the dual space of $C(X)$. Of course, the value of $f \in C(X)$ at $\mu \in S$ is given by $\mu(f) = \int_X f\mathrm{d}\mu$ It is not hard to see that S is a simplex and, in fact, it is even a regular (Bauer in [1]) simplex. Moreover, identifying each $x \in X$ with the point measure $\delta_x \in S$, the space X becomes (homeomorphic to) the set ∂S of all extreme points of S.

Now, observe that the diameter of the range of f viewed as an element of $C(X)$ equals the diameter of the range of f viewed as an affine map on S. That $\varrho_X(f) \leq \varrho_S(f)$ is obvious. As for the reverse inequality, fix $f \in C(X)$ and define a mapping $g : S \times S \to \mathbb{K}$ as $g(\mu, \lambda) = \mu(f) - \lambda(f)$. Clearly, $S \times S$ is a compact convex set and g is affine and continuous. By the maximum principle for affine functions ([9], Proposition 23.1.10) the maximum value of $|g|$ is attained at some extreme point of $S \times S$. Since $\partial(S \times S) = \partial S \times \partial S$, it follows that

$$\varrho_S(f) = \sup_{\mu,\lambda \in S} |g(\mu, \lambda)| = \max_{x,y \in X} |g(\delta_x, \delta_y)| = \max_{x,y \in X} |f(x) - f(y)| = \varrho_X(f),$$

as desired.

On the other hand, each affine automorphism φ of S induces by restriction a homeomorphism ϕ of $X = \partial S$. (Actually, this restriction determines φ, by the Krein–Milman theorem. Conversely, if ϕ is a homeomorphism of X, then the map $\varphi : S \to S$ given by $(\varphi(\mu))(A) = \mu(\phi^{-1}(A))$ is the unique affine mapping whose restriction to X coincides with ϕ.) Clearly, $f \circ \phi \in C(X)$ and $f \circ \varphi \in \mathcal{A}(S)$ coincide via the identification between $C(X)$ and $\mathcal{A}(S)$ described above. It is now apparent that Theorem 2 is nothing but a particular case of Theorem 1.

Returning to affine functions, let $\mathcal{A}_\varrho(S)$ denote the quotient of the space $\mathcal{A}(S)$ by the kernel of ϱ. Clearly, $\mathcal{A}_\varrho(S)$ is a Banach space under the norm

$$\|\pi(f)\|_\varrho = \varrho(f),$$

where $\pi : \mathcal{A}(S) \to \mathcal{A}(S)/\mathrm{ker}\varrho$ is the natural quotient map.

Suppose that $T : \mathcal{A}(S) \to \mathcal{A}(S)$ is a diameter preserving linear bijection. Then there exists a (unique) isometry T_ϱ of $\mathcal{A}_\varrho(S)$ making commute the following diagram

$$\begin{tikzcd}
\mathcal{A}(S) \arrow[r, "T"] \arrow[d, "\pi"'] & \mathcal{A}(S) \arrow[d, "\pi"] \\
\mathcal{A}_\varrho(S) \arrow[r, "T_\varrho"] & \mathcal{A}_\varrho(S)
\end{tikzcd}.$$

The main step in the proof of Theorem 1 is the following characterization of the isometries of $\mathcal{A}_\varrho(S)$.

Theorem 3. *Let S be a simplex. A linear map $T_\varrho : \mathcal{A}_\varrho(S) \to \mathcal{A}_\varrho(S)$ is a surjective isometry if and only if there is an affine automorphism φ of S and $\tau \in \mathbb{K}$, with $|\tau| = 1$, such that $T_\varrho(\pi(f)) = \pi(\tau f \circ \varphi)$, for all $f \in \mathcal{A}(S)$.*

Remark 1. Our main result has been independently obtained, in a more general setting, by Rao and Roy in [8]. This paper contains other interesting results about function algebras and vector-valued maps.

2. Proofs

We shall keep the organisation of [2]. First, we derive Theorem 1 from Theorem 3.

Proof of Theorem 1. Let T be a diameter preserving bijection of $\mathcal{A}(S)$ and let $T_\varrho : \mathcal{A}_\varrho(S) \longrightarrow \mathcal{A}_\varrho(S)$ be the corresponding isometry. According to Theorem 3, one has $T_\varrho(\pi(f)) = \pi(\tau f \circ \varphi)$, for suitable φ and τ. Since $T_\varrho \circ \pi = \pi \circ T$ one has

$$\pi(Tf) = \pi(\tau f \circ \varphi),$$

so that $f \longrightarrow Tf - \tau f \circ \varphi$ takes values in the subspace of constant functions of $\mathcal{A}(S)$ (which is the kernel of π). This obviously implies that there is $\mu : \mathcal{A}(S) \longrightarrow \mathbb{K}$ such that

$$Tf = \tau f \circ \varphi + \mu(f) 1_S$$

for every $f \in \mathcal{A}(S)$. □

Remark 2. Observe that T need not be continuous. In fact, T is continuous if and only if μ is.

For the proof of Theorem 3 we need a description of the extreme points of U^*, the unit ball of $\mathcal{A}_\varrho(S)^*$.

Lemma 1. *Let $\mu \in \mathcal{A}_\varrho(S)^*$. Then μ is an extreme point of U^* if and only if $\mu = \sigma(\delta_x - \delta_y)$, where x and y are distinct extreme points of S and $|\sigma| = 1$.*

Proof. First note that for any compact convex set K the extreme points of the unit ball of $\mathcal{A}(K)^*$ (the dual space of $\mathcal{A}(K)$ which is equipped with the usual supremum norm $\|f\|_\infty = \sup\{|f(x)| : x \in K\}$ unless otherwise stated) have the form $\sigma\delta_e$ for some $e \in \partial K$ and $|\sigma| = 1$.

Necessity. Consider the linear operator $L : \mathcal{A}_\varrho(S) \longrightarrow \mathcal{A}(S \times S)$ given by

$$L(\pi(f))(x, y) = f(x) - f(y).$$

Obviously, $\|\pi(f)\|_\varrho = \|Lf\|_\infty$, so that L is an isometric embedding. The Hahn–Banach theorem implies that μ is an extreme point of U^*, then μ is the restriction to $\mathcal{A}_\varrho(S)$ of some extreme point on the ball of $\mathcal{A}(S \times S)^*$, so $\mu = L^*(\sigma\delta_{(x,y)})$, where $|\sigma| = 1$ and $(x, y) \in \partial(S \times S)$ (that is $x, y \in \partial S$). Hence,

$$\mu = \sigma L^* \delta_{(x,y)} = \sigma(\delta_x - \delta_y)$$

with $x \neq y$. This proves the ‘only if’ part.

Sufficiency. Let us assume for a moment that $\mathbb{K} = \mathbb{R}$. One then has

$$\varrho(f) = 2 \inf\{\|f - \lambda 1_S\|_\infty : \lambda \in \mathbb{R}\}$$

for all $f \in \mathcal{A}(S)$. This means that $\mathcal{A}_\varrho(S)$ is, up to a constant factor 2, isometric to the quotient of $(\mathcal{A}(S), \|\cdot\|_\infty)$ by the subspace of constant functions (which is not true for

$\mathbb{K} = \mathbb{C}$). Therefore, the space $\mathcal{A}_\varrho(S)^*$ is, up to a factor 1/2, isometric (and not only isomorphic) to a subspace of $\mathcal{A}(S)^*$.

In fact, we have $\mathcal{A}_\varrho(S)^* = \{\mu \in \mathcal{A}(S)^* : \mu(1_S) = 0\}$, with

$$2\|\mu\|_{\mathcal{A}_\varrho(S)^*} = \|\mu\|_{\mathcal{A}(S)^*}.$$

So, one can work with $\|\cdot\|_{\mathcal{A}(S)^*}$ instead of the original norm of $\mathcal{A}_\varrho(S)^*$. This is the point where the hypothesis that S is a simplex (and not merely a compact convex set) appears.

Being S a simplex, the dual space $(\mathcal{A}(S)^*, \|\cdot\|_{\mathcal{A}(S)^*})$ which is always an ordered space (define $\mu \in \mathcal{A}(S)^*$ to be non-negative provided $\mu(f) \geq 0$ for every non-negative $f \in \mathcal{A}(S)$) becomes a Banach lattice. In fact, $\mathcal{A}(S)^*$ is an abstract L-space (that is, a Banach lattice where the norm is additive in the positive cone) and, by an old result of Kakutani, it is order isometric to some concrete L_1-space (see [9]). In particular, if λ^+ and λ^- denote respectively the positive and negative part of $\lambda \in \mathcal{A}(S)^*$, then

$$\|\lambda\|_{\mathcal{A}(S)^*} = \|\lambda^+\|_{\mathcal{A}(S)^*} + \|\lambda^-\|_{\mathcal{A}(S)^*}.$$

Moreover, it is easily seen that if $\lambda = \lambda_1 - \lambda_2$ is a decomposition of λ with λ_1 and λ_2 non-negative and $\|\lambda\|_{\mathcal{A}(S)^*} = \|\lambda_1\|_{\mathcal{A}(S)^*} + \|\lambda_2\|_{\mathcal{A}(S)^*}$, then $\lambda^+ = \lambda_1$ and $\lambda^- = \lambda_2$. Also note that $\mu \in \mathcal{A}(S)^*$ is non-negative if and only if $\|\mu\|_{\mathcal{A}(S)^*} = \mu(1_S)$.

After these preparatives, let $x, y \in \partial S$ with $x \neq y$. Note that $\|\delta_x - \delta_y\|_{\mathcal{A}(S)^*} = 2$ since δ_x and δ_y are distinct extreme points in the unit ball of an abstract L-space. (Actually it is not hard to find a continuous affine map $f : S \to [-1, 1]$ such that $f(x) = 1$ and $f(y) = -1$.) Thus, $\delta_x = (\delta_x - \delta_y)^+$ and $\delta_y = (\delta_x - \delta_y)^-$. Now, suppose $\mu, \nu \in \mathcal{A}_\varrho(S)^*$ are such that $\delta_x - \delta_y = \mu + \nu$, with $\|\delta_x - \delta_y\|_{\mathcal{A}_\varrho(S)^*} = \|\mu\|_{\mathcal{A}_\varrho(S)^*} + \|\nu\|_{\mathcal{A}_\varrho(S)^*}$. Writing $\mu = \mu^+ - \mu^-$ and $\nu = \nu^+ - \nu^-$, one obtains

$$\begin{aligned}\|\delta_x - \delta_y\|_{\mathcal{A}(S)^*} &= \|\mu\|_{\mathcal{A}(S)^*} + \|\nu\|_{\mathcal{A}(S)^*} \\ &= \|\mu^+\|_{\mathcal{A}(S)^*} + \|\mu^-\|_{\mathcal{A}(S)^*} + \|\nu^+\|_{\mathcal{A}(S)^*} + \|\nu^-\|_{\mathcal{A}(S)^*} \\ &= \|\mu^+ + \nu^+\|_{\mathcal{A}(S)^*} + \|\mu^- + \nu^-\|_{\mathcal{A}(S)^*}.\end{aligned}$$

It follows that $\delta_x = (\delta_x - \delta_y)^+ = \mu^+ + \nu^+$ and $\delta_y = (\delta_x - \delta_y)^- = \mu^- + \nu^-$. Hence $\|\delta_x\|_{\mathcal{A}(S)^*} = \|\mu^+\|_{\mathcal{A}(S)^*} + \|\nu^+\|_{\mathcal{A}(S)^*}$ and $\|\delta_y\|_{\mathcal{A}(S)^*} = \|\mu^-\|_{\mathcal{A}(S)^*} + \|\nu^-\|_{\mathcal{A}(S)^*}$. Since δ_x and δ_y are extreme points in the unit ball of $\mathcal{A}(S)^*$, one obtains

$$\begin{aligned}\mu^+ &= \mu^+(1_S)\delta_x, \quad & \nu^+ &= \nu^+(1_S)\delta_x \\ \mu^- &= \mu^-(1_S)\delta_y, \quad & \nu^- &= \nu^-(1_S)\delta_y.\end{aligned}$$

But μ and ν belong to $\mathcal{A}_\varrho(S)^*$, so we have $\mu^+(1_S) = \mu^-(1_S)$ and $\nu^+(1_S) = \nu^-(1_S)$ and, therefore,

$$\begin{aligned}\mu &= \mu^+(1_S)(\delta_x - \delta_y), \\ \nu &= \nu^+(1_S)(\delta_x - \delta_y).\end{aligned}$$

This shows that $\delta_x - \delta_y$ is an extreme point of U^* in the real case.

To end with the proof of the Lemma, let $\mathbb{K} = \mathbb{C}$. It obviously suffices to see that $\delta_x - \delta_y$ is an extreme point of the unit ball of the complex $\mathcal{A}_\varrho(S)^*$. Suppose that

$$\delta_x - \delta_y = \mu + \nu \quad \text{and} \quad \|\delta_x - \delta_y\|_{\mathcal{A}_\varrho(S)^*} = \|\mu\|_{\mathcal{A}_\varrho(S)^*} + \|\nu\|_{\mathcal{A}_\varrho(S)^*}.$$

Thinking $\mathcal{A}_\varrho(S)$ as a subspace of $C(S \times S)$ via the map L constructed above, the Hahn–Banach theorem implies the existence of extensions $\tilde{\mu}, \tilde{\nu} \in M(S \times S)$ so that

$$T^*\tilde{\mu} = \mu \quad \text{with} \quad \|\tilde{\mu}\|_1 = \|\mu\|_{\mathcal{A}_\varrho(S)^*}$$
$$T^*\tilde{\nu} = \nu \quad \text{with} \quad \|\tilde{\nu}\|_1 = \|\nu\|_{\mathcal{A}_\varrho(S)^*}.$$

Since $\|\Re(\eta)\|_1 \leq \|\eta\|_1$ for every $\eta \in M(S \times S)$, with equality only if η is real, it follows that $\tilde{\mu}$ and $\tilde{\nu}$ are real measures. Hence $\mu(\pi(f))$ and $\nu(\pi(f))$ are real for every real-valued $f \in \mathcal{A}(S)$ and

$$(\delta_x - \delta_y)|_{\mathcal{A}_\varrho(S,\mathbb{R})} = \mu|_{\mathcal{A}_\varrho(S,\mathbb{R})} + \nu|_{\mathcal{A}_\varrho(S,\mathbb{R})}$$

with

$$\|(\delta_x - \delta_y)|_{\mathcal{A}_\varrho(S,\mathbb{R})}\|_{\mathcal{A}_\varrho(S,\mathbb{R})^*} = \|\mu|_{\mathcal{A}_\varrho(S,\mathbb{R})}\|_{\mathcal{A}_\varrho(S,\mathbb{R})^*} + \|\nu|_{\mathcal{A}_\varrho(S,\mathbb{R})}\|_{\mathcal{A}_\varrho(S,\mathbb{R})^*},$$

since obviously $\|(\delta_x - \delta_y)|_{\mathcal{A}_\varrho(S,\mathbb{R})}\|_{\mathcal{A}_\varrho(S,\mathbb{R})^*} = 2$. On the other hand $(\delta_x - \delta_y)|_{\mathcal{A}_\varrho(S,\mathbb{R})}$ is an extreme point of the unit ball of $\mathcal{A}_\varrho(S, \mathbb{R})^*$ and, therefore, μ and ν are proportional to $\delta_x - \delta_y$ when restricted to real functions. By complex linearity one obtains that μ and ν also are proportional to $\delta_x - \delta_y$ as complex functionals. This completes the proof of Lemma 1. □

Beginning of the proof of Theorem 3. Let T be a surjective isometry of $\mathcal{A}_\varrho(S)$. Then the adjoint map $T^* : \mathcal{A}_\varrho(S)^* \longrightarrow \mathcal{A}_\varrho(S)^*$ is an isometry as well and, therefore, it sends the set of extreme points of U^* into itself. Taking Lemma 1 into account, it is clear that, given $x, y \in \partial S$ with $x \neq y$, there are $u, v \in \partial S$, $u \neq v$ and $\sigma \in \mathbb{K}$ with $|\sigma| = 1$ such that

$$T^*(\delta_x - \delta_y) = \sigma(\delta_u - \delta_v).$$

Let ∂S_2 stand for the collection of all subsets of ∂S having exactly two elements. Plainly, T induces a bijection $\partial S_2 \longrightarrow \partial S_2$ by

$$\Phi\{x, y\} = \operatorname{supp}(T^*(\delta_x - \delta_y)).$$

The definition of Φ makes sense because if x, y, z, w are extreme points of S with $x \neq y$, $z \neq w$ and $\delta_x - \delta_y = \sigma(\delta_z - \delta_w)$ for some unimodular σ, then $\{x, y\} = \{z, w\}$. This follows from the fact that if e, f, g, h are positive extreme points of the unit ball of an abstract L-space with $e \neq f, g \neq h$ and $e - f = h - g$, then $e = h$ and $f = g$.

Let $|A|$ denote the cardinality of the set A.

Lemma 2. *For all* $\{x, y\}, \{u, v\} \in \partial S_2$, *one has* $|\{x, y\} \cap \{u, v\}| = |\Phi\{x, y\} \cap \Phi\{u, v\}|$.

Proof of Lemma 2. Simply observe that if $\{x, y\} \neq \{u, v\}$, then $\{x, y\} \cap \{u, v\}$ is nonempty if and only if there is a nontrivial linear combination of $\delta_x - \delta_y$ and $\delta_u - \delta_v$ that is an extreme point of the unit ball of $\mathcal{A}_\varrho(S)^*$. □

Lemma 3. (see [2], Lemma 3). *There is a bijection* $\phi : \partial S \longrightarrow \partial S$ *such that* $\Phi\{x, y\} = \{\phi(x), \phi(y)\}$ *for every* $x, y \in \partial S$. □

End of the proof of Theorem 3. Let $\phi : \partial S \longrightarrow \partial S$ be the (obviously bijective) map of the preceding lemma. Clearly

$$T^*(\delta_x - \delta_y) = \sigma(x, y)(\delta_{\phi(x)} - \delta_{\phi(y)}),$$

where $|\sigma(x,y)| = 1$. We want to see that $\sigma(x,y)$ does not depend on x,y. Let $z \notin \{x,y\}$. Then

$$\begin{aligned}\sigma(x,y)(\delta_{\phi(x)} - \delta_{\phi(y)}) &= T^*(\delta_x - \delta_y)\\ &= T^*(\delta_x - \delta_z + \delta_z - \delta_y)\\ &= T^*(\delta_x - \delta_z) + T^*(\delta_z - \delta_y)\\ &= \sigma(x,z)(\delta_{\phi(x)} - \delta_{\phi(z)}) + \sigma(z,y)(\delta_{\phi(z)} - \delta_{\phi(y)}),\end{aligned}$$

so that

$$\sigma(x,y) = \sigma(x,z) = \sigma(z,y).$$

Since x,y and z are arbitrary, the equality $\sigma(x,y) = \sigma(z,y)$ means that $\sigma(\cdot,\cdot)$ does not depend on the first variable, while $\sigma(x,y) = \sigma(x,z)$ implies that the same occurs with the second one. Hence $\sigma(x,y) = \tau$ for some unimodular τ.

Our next objective is to extend $\phi : \partial S \to \partial S$ to an automorphism ψ of the simplex S. (Notice that if S were a regular simplex, that is, with closed extreme boundary, this would be automatic.)

Without loss of generality, assume that $\tau = 1$. Fix $y \in \partial S$ and define an affine mapping $\psi : S \to A_\varrho(S)^*$ by

$$\psi(x) = T^*(\delta_x - \delta_y) + \delta_{\phi(y)}.$$

Observe that ψ is continuous when $A_\varrho(S)^*$ is endowed with the weak* topology. Since $\psi(x) = \delta_{\phi(x)}$ for every $x \in \partial S$, it follows that ψ takes values in S (that is, in the canonical image of S in $A_\varrho(S)^*$).

Thus, we can define a continuous affine map $\varphi : S \to S$ by the condition

$$\delta_{\varphi(x)} - \delta_{\psi(y)} = T^*(\delta_x - \delta_y).$$

Notice that φ is in fact an affine automorphism (whose inverse can be obtained from T^{-1}). Finally, define $T_{(\tau,\varphi)} : A_\varrho(S) \to A_\varrho(S)$ as $T_{(\tau,\varphi)}(\pi(f)) = \pi(\tau f \circ \varphi)$. Since

$$T^*(\delta_x - \delta_y) = T^*_{(\tau,\varphi)}(\delta_x - \delta_y)$$

for all $x,y \in \partial S$, the Krein–Milman theorem implies that $T = T_{(\tau,\varphi)}$. This completes the proof of Theorem 3. □

Acknowledgement

This work is supported in part by DGICYT project PB97-0377.

References

[1] Alfsen E M, Compact convex sets and boundary integrals, *Ergebnisse der Math.* **57** (Springer) (1971)

[2] Cabello Sánchez F, Diameter preserving linear maps and isometries, *Arch. Math. (Basel)* **73** (1999) 373–379

[3] González F and Uspenskij V V, On homomorphisms of groups of integer-valued functions, *Extracta Math.* **14** (1999) 19–29

[4] Györy M and Molnár L, Diameter preserving linear bijections of $C(X)$, *Arch. Math. (Basel)* **71** (1998) 301–310

[5] Li C-K and Tsing N-K, Linear preserver problems: a brief introduction and some special techniques, *Linear Algebra Appl.* **162–164** (1992) 217–235

[6] Lindenstrauss J, Some useful facts about Banach spaces, (Springer Lecture Notes in Mathematics) 1317

[7] Poulsen E T, A simplex with dense extreme boundary, *Ann. Inst. Fourier* **11** (1961) 83–87

[8] Rao T S S R K and Roy A K, Diameter preserving linear bijections of function spaces (preprint 1999)

[9] Semadeni Z, Banach Spaces of Continuous Functions, *Monografie Matematyczne* **55** (1971) PWN

Proc. Indian Acad. Sci. (Math. Sci.), Vol. 110, No. 2, May 2000, pp. 213–223.
© Printed in India

On vector equilibrium problem

K R KAZMI

Department of Mathematics, Aligarh Muslim University, Aligarh 202 002, India

MS received 22 October 1996; revised 24 January 2000

Abstract. This paper presents some existence results of a vector equilibrium problem. The several important special cases of the vector equilibrium problem are also discussed.

Keywords. Vector equilibrium problem; KKM–Fan lemma; P-monotone and P-convex mappings; P-maximum monotonicity.

1. Introduction

Let X be a real topological vector space; $K \subset X$ be a nonempty closed convex set; (Y, P) be a real ordered topological vector space with a partial order (or vector order) $\leq_P$ induced by a solid pointed closed convex cone P, thus $x \leq_P y \Longleftrightarrow y - x \in P$, $\forall x, y \in Y$; $f : X \times X \longrightarrow Y$ with $f(x,x) = 0$ for all $x \in X$. The vector equilibrium problem is to find $x \in K$, such that

$$\text{(VEP)} \qquad f(x,y) \notin -\operatorname{int} P, \quad \forall y \in K,$$

where $\operatorname{int} P$ denotes interior of P. This problem includes as special cases, vector optimization problems, vector complementarity problems, fixed points problems, vector variational inequality problems etc. If $Y = \mathbb{R}$, $P = \mathbb{R}_+$ then (VEP) reduces to the equilibrium problem of finding $x \in K$ such that

$$f(x,y) \geq 0, \quad \forall y \in K. \tag{1}$$

This problem was considered and studied by Blum and Oettli [B–R]. In this paper, by making use of KKM–Fan lemma [F1], we prove some existence results for the vector equilibrium problem (VEP) in the case where

$$f(x,y) = g(x,y) + f(x,y). \tag{2}$$

Also, we review some special cases of (VEP).

2. Special cases

In this section we review some of the important examples for vector equilibrium problem (VEP). In the examples below X^*, the topological dual of X, should be topologized in such a way that the cannonical bilinear form $\langle \cdot, \cdot \rangle$ is continuous on $X^* \times X^*$.

DEFINITION 1

A function $f(\cdot,\cdot) : K \times K \longrightarrow Y$ is called P-monotone if and only if

$$f(x,y) + f(y,x) \in -P, \quad \forall x, y \in K.$$

(i) *Vector optimization*

Let $\phi : K \longrightarrow Y$, then to find $x \in K$ such that

$$\phi(y) - \phi(x) \notin -\operatorname{int} P, \quad \forall y \in K. \tag{3}$$

Equation (3) is equivalent to the following:

$$W_{\min}\phi(x) \text{ subject to } x \in K, \tag{4}$$

where $W_{\min}$ denotes weak minima, see for instance [C] and [C–C2]. Setting

$$f(x,y) := \phi(y) - \phi(x).$$

Then problem (3) coincides with (VEP) provided that f is P-monotone.

(ii) *Convex differentiable vector optimization*

Besides the significant connection between vector optima and vector equilibrium given in preceding example, there is a more subtle connection in the convex and differentiable case. Let $\phi : X \longrightarrow Y$ be P-convex and linear Gateaux differentiable. Then the problem (4) and the vector variational inequality of finding

$$x \in K, \quad \text{such that } \langle \phi'(x), y - x \rangle \notin -\operatorname{int} P, \quad \forall y \in K, \tag{5}$$

have the same set of solutions, see for instance [C] and [C–C2].

Upon setting $f(x,y) := \langle \phi'(x), y - x \rangle$ this becomes an example of (VEP). The function f is P-monotone in this case since $\phi'(\cdot)$ is P-monotone.

(iii) *Nonconvex differentiable vector optimization*

We can replace convexity by invexity in the preceding example as follows:

DEFINITION 2

A set $K \subseteq X$ is called *invex* if there is a mapping $\eta : K \times K \longrightarrow K$ such that, for every $x, y \in K$ and $\lambda \in [0,1]$, there holds $y + \lambda\eta(x,y) \in K$.

DEFINITION 3

Let K be an invex set then $\phi : K \longrightarrow Y$ is called *P-preinvex* if, for every $x, y \in K$ and $\lambda \in [0,1]$,

$$\lambda\phi(y) - (1-\lambda)\phi(x) - \phi(x + \lambda\eta(y,x)) \in P, \quad \text{see [W–J]}.$$

Now, if $\phi : K \longrightarrow Y$ be P-preinvex and Fréchet (or linear Gateaux) differentiable then problem (4) and the vector variational-like inequality problem of finding $x \in K$ such that

$$\langle \phi'(x), \eta(y,x) \rangle \notin -\operatorname{int} P, \quad \forall y \in K \tag{6}$$

have the same set of solutions, see for instance [K2]. For related work, see [K1, K3, K–A].

Upon setting $f(x,y) := \langle \phi'(x), \eta(y,x) \rangle$ this becomes an example of (VEP). The function f is $P - \eta$-monotone in the case, since the mapping $\phi'(\cdot)$ is $P - \eta$-monotone with $\eta(x,y) = -\eta(y,x)$.

(iv) *Vector variational inequalities*

$L(X,Y)$ denotes the space of all linear bounded operators from X into Y. Let $T : K \longrightarrow L(X,Y)$, find $x \in X$ such that

$$x \in K, \quad \langle Tx, y-x \rangle \notin -\operatorname{int} P \quad y \in K. \tag{7}$$

Vector variational inequalities were first introduced by [G]. Set $f(x,y) := \langle Tx, y-x \rangle$. Then (7) $\Longleftrightarrow$ (VEP).

(v) *Vector complementarity problems*

This is special case of the previous example. Let K be a closed convex cone in X. The weak P-dual cone $K_P^{w^+}$ of K is defined by

$$K_P^{w^+} = \{l \in L(X,Y) : \langle l, x \rangle \notin -\operatorname{int} P, \quad \forall x \in K\}.$$

The strong P-dual cone $K_P^{s^+}$ of K is defined by

$$K_P^{s^+} = \{l \in L(X,Y) : \langle l, x \rangle \in P, \quad \forall x \in K\}.$$

Let $T : X \longrightarrow L(X,Y)$ be a given mapping. Then the vector complementarity problems:

$$\text{Find } x \in X \quad \text{such that} \quad x \in K, \quad Tx \in K_P^{w^+}, \quad \langle Tx, x \rangle \notin \operatorname{int} P \tag{8}$$

and

$$\text{Find } x \in X \quad \text{such that} \quad x \in K, \quad Tx \in K_P^{s^+}, \quad \langle Tx, x \rangle \notin \operatorname{int} P. \tag{9}$$

Problem (9) $\Rightarrow$ problem (7) $\Rightarrow$ problem (8), see [Y]. But we have seen that problem (7) is equivalent to (VEP).

(vi) *Fixed-point problem*

For each $x \in K$ let

$$F(x) := \{z \in K : \langle T(x), y-z \rangle \notin -\operatorname{int} P, \quad \forall y \in K\}.$$

Then the fixed point problem:

$$\text{Find } x \in K \text{ such that } x \in F(x). \tag{10}$$

Problem (10) $\Longleftrightarrow$ problem (7), see for instance [Y].

3. Existence results

In this section, we prove some existence results for (VEP) in the case where

$$f(x,y) = g(x,y) + h(x,y).$$

We need the following definitions and result.

DEFINITION 4

Let K and C be convex sets with $C \subset K$. Then $\operatorname{core}_K C$, the core of C relative to K, is defined as

$$a \in \operatorname{core}_K C \Longleftrightarrow (a \in C, \text{ and } C \cap (a,y) \neq 0 \quad \forall y \in K \setminus C).$$

Note that $\operatorname{core}_K K = K$.

DEFINITION 5

Let (Y, P) be an ordered topological vector space. $T : X \longrightarrow Y$ is called *P-convex* iff for each pair $x, y \in X$ and $\lambda \in [0, 1]$,

$$T(\lambda y + (1 - \lambda)x) \leq_P \lambda T(y) + (1 - \lambda)T(x).$$

Lemma 6. *See* [C]. *Let (Y, P) be an ordered topological vector space with a solid pointed closed convex cone P. Then, $\forall x,\ y \in X$, we have*

(i) $y - x \in \operatorname{int} P$ *and* $y \notin \operatorname{int} P$ imply $x \notin \operatorname{int} P$;
(ii) $y - x \in P$ *and* $y \notin \operatorname{int} P$ imply $x \notin \operatorname{int} P$;
(iii) $y - x \in -\operatorname{int} P$ *and* $y \notin -\operatorname{int} P$ imply $x \notin -\operatorname{int} P$;
(iv) $y - x \in -P$ *and* $y \notin -\operatorname{int} P$ imply $x \notin -\operatorname{int} P$.

Theorem 7. *Let the following assumptions hold:*

(i) *X is a real topological vector space; $K \subset X$ is a closed convex nonempty set; (Y, P) is a real ordered topological vector space with a solid pointed closed convex cone P in Y.*
(ii) *$g : X \times X \longrightarrow Y$ has the following properties: $g(x, x) = 0$, $\forall x \in K$; g is P-monotone; $\forall x, y \in K$, the function $t \in [0, 1] \longrightarrow g(ty + (1 - t)x, y)$ is continuous at 0_+; g is a P-convex and continuous in the second argument.*
(iii) *$h : X \times X \longrightarrow Y$ has the following properties: $h(x, x) = 0$, $\forall x \in K$; h is continuous in the first argument; h is P-convex in the second argument.*
(iv) *There exists a nonempty compact convex C of K such that for every $x \in C \backslash \operatorname{core}_K C$ there exists $a \in \operatorname{core}_K C$ such that*

$$g(x, a) + h(x, a) \in -P.$$

Then there exists $x \in C$ such that

$$g(x, y) + h(x, y) \notin -\operatorname{int} P, \quad \forall y \in K.$$

First we shall prove the following three lemmas, for which the hypotheses remain the same as for theorem 7.

Lemma 8. *There exists $x \in C$ such that*

$$h(x, y) - g(y, x) \notin -\operatorname{int} P, \quad \forall y \in C.$$

Proof. Let, for each fixed $y \in C$,

$$S(y) := \{x \in C : h(x, y) - g(y, x) \notin -\operatorname{int} P\}.$$

Claim that $\cap_{y \in C} S(y) \neq \phi$. Indeed, let $\{y_1, \ldots, y_n\}$ be a finite subset of C. Let $I \subset N$ be nonempty; let $z \in \operatorname{conv}\{y_i : i \in I\}$ be arbitrary. Then $z = \sum_{i \in I} \lambda_i y_i$ with $\lambda_i \geq 0$ and $\sum_{i \in I} \lambda_i = 1$. Suppose, if possible, $z \notin \cup_{i \in I} S(y_i)$. Then

$$h(z, y_i) - g(y_i, z) \in -\operatorname{int} P, \quad \forall i \in I.$$

From this follows

$$\sum_{i \in I} \lambda_i h(z, y_i) - \sum_{i \in I} \lambda_i g(y_i, z) \in -\operatorname{int} P. \tag{11}$$

Now, since g is P-convex and p-monotone, then we have

$$\begin{aligned}\sum_{i\in I}\lambda_i h(z,y_i) &\leq_P \sum_{i\in I}\sum_{j\in I}\lambda_i\lambda_j g(y_i,y_j)\\ &=\frac{1}{2}\sum_{i,j\in I}\lambda_i\lambda_j(g(y_i,y_j)+g(y_i,y_j))\\ &\leq_P 0,\end{aligned}\tag{12}$$

and from the properties of h follows

$$0=h(z,z)\leq_P\sum_{i\in I}\lambda_i h(z,y_i).\tag{13}$$

From (12) and (13) and, using the properties of P, we have

$$\sum_{i\in I}\lambda_i h(z,y_i)-\sum_{i\in I}\lambda_i g(y_i,z)\in P,$$

which is a contradiction to (11). Hence, our supposition is false. Thus

$$\operatorname{conv}\{y_i : i\in I\}\subset\bigcup_{i\in I}S(y_i).$$

Also, this is true for every nonempty subset I of N. The sets $S(y_i)$ are closed for every i. Indeed, let $\{x_n^i\}$ be a sequence in $S(y_i)$ such that $x_n^i\longrightarrow x^i$ then

$$\begin{aligned}&h(x_n^i,y_i)-g(y_i,x_n^i)\notin-\operatorname{int}P\\ &\Longrightarrow h(x_n^i,y_i)-g(y_i,x_n^i)\in W:=Y\backslash(-\operatorname{int}P).\end{aligned}$$

Since W is closed and h and g are continuous in the first and second argument respectively, then we have

$$h(x^i,y_i)-g(y_i,x^i)\in W.$$

This implies that the sets $S(y_i)$ are closed for every i. Hence it follows from the KKM–Fan lemma that

$$\cap_{i\in N}S(y_i)\neq\phi.$$

In other words, any finite subfamily $S(y)$ for each $y\in C$, has nonempty intersection. Since these sets are closed subsets of the compact set C, it follows that the entire family has nonempty intersection.

Hence

$$\cap_{y\in C}S(y)\neq\phi,$$

i.e. there exists atleast one $x\in C$ such that

$$h(x,y)-g(y,x)\notin-\operatorname{int}P,\quad\forall y\in C.$$

Lemma 9. *The following statements are equivalent:*

(A) $x\in C,\quad h(x,y)-g(y,x)\notin-\operatorname{int}P,\quad\forall y\in C;$

(B) $x\in C,\quad h(x,y)+g(y,x)\notin-\operatorname{int}P,\quad\forall y\in C.$

Proof. Let (B) hold. Since g is P-monotone,

$$g(x,y) \leq_P -g(y,x).$$

Also,

$$h(x,y) + g(x,y) \leq_P h(x,y) - g(y,x). \tag{14}$$

Since $h(x,y) + g(x,y) \notin - \operatorname{int} P$, using (iv) of lemma 6, (14) implies (A).

Let (A) hold. Let $y \in C$ be arbitrary, and let $x_\lambda := \lambda y + (1-\lambda)x$, $0 < \lambda \leq 1$. Then $x_\lambda \in C$, and hence from (A)

$$(1-\lambda)h(x,x_\lambda) - (1-\lambda)g(x_\lambda,x) \notin - \operatorname{int} P. \tag{15}$$

Since g is P-convex in the second argument and $g(x,x) = 0$, $\forall x \in C$ then for all $0 < \lambda \leq 1$,

$$0 = g(x_\lambda,x_\lambda) \leq_P \lambda g(x_\lambda,y) + (1-\lambda)g(x_\lambda,x)$$

or,

$$-(1-\lambda)g(x_\lambda,x) \leq_P \lambda g(x_\lambda,y).$$

Since $(1-\lambda)h(x,x_\lambda) \in Y$,

$$(1-\lambda)h(x,x_\lambda) - (1-\lambda)g(x_\lambda,x) \leq_P (1-\lambda)h(x,x_\lambda) + \lambda g(x_\lambda,y). \tag{16}$$

From (15) and (16) and using (iv) of lemma 6, we have

$$(1-\lambda)h(x,x_\lambda) + \lambda g(x_\lambda,y) \notin - \operatorname{int} P. \tag{17}$$

Since h is P-convex in the second argument and $h(x,x) = 0$, $\forall x \in C$, then

$$\lambda(1-\lambda)h(x,y) - \lambda g(x_\lambda,y)\, {}_P\!\geq (1-\lambda)h(x,x_\lambda) + \lambda g(x_\lambda,y) \notin - \operatorname{int} P$$

and hence by (iv) of lemma 6,

$$\lambda(1-\lambda)h(x,y) + \lambda g(x_\lambda,y) \notin - \operatorname{int} P.$$

Dividing by $\lambda > 0$ we obtain

$$g(x_\lambda,y) + (1-\lambda)h(x,y) \notin - \operatorname{int} P.$$

Since g is hemicontinuous in the first argument, it follows that

$$g(x,y) + h(x,y) \in -\operatorname{int} P$$

as $\lambda \longrightarrow 0_+$. Hence (B) holds.

Lemma 10. *Assume that* $\phi : K \longrightarrow Y$ *is* P-*convex,* $x_0 \in \operatorname{core}_K C, \phi(x_0) \notin \operatorname{int} P$, *and* $\phi(y) \notin \operatorname{int} P$, $\forall y \in C$. *Then* $\phi(y) \notin - \operatorname{int} P$, $\forall y \in K$.

Proof of lemma 10 is omitted because it can be obtained by using the same arguments as used in the proof of lemma 4 of [B–R].

Proof of theorem 7. By lemma 8, it follows that there exists atleast one $x \in C$ such that

$$h(x,y) - g(y,x) \notin - \operatorname{int} P, \quad \forall y \in C.$$

By lemma 9, it follows that the above assertion is equivalent to

$$h(x,y) - g(x,y) \notin -\operatorname{int} P, \quad \forall y \in C.$$

Set

$$\phi(\cdot) := h(x,\cdot) + g(x,\cdot).$$

Clearly $\phi(\cdot)$ is P-convex and

$$\phi(y) \notin -\operatorname{int} P, \quad \forall y \in C.$$

If $x \in \operatorname{core}_K C$, then set $x_0 := x$. If $x \in C \setminus \operatorname{core}_K C$, then set $x_0 := a$, where a is as in assumption (iv). In both cases $x_0 \in \operatorname{core}_K C$, and $\phi(x_0) \notin \operatorname{int} P$. Hence, by lemma 10, it follows that

$$\phi(y) \notin -\operatorname{int} P, \quad \forall y \in K,$$

i.e.

$$g(x,y) + h(x,y) \notin -\operatorname{int} P, \quad \forall y \in K.$$

Thus, there exists atleast one $x \in C$ such that

$$g(x,y) + h(x,y) \notin -\operatorname{int} P, \quad \forall y \in K.$$

Let $Y = \mathbb{R}, P = \mathbb{R}_+$. If $g = 0$ then theorem 7 becomes a variant of Ky Fan's minimax theorem [F2], whereas for $h = 0$ it becomes a variant of the Browder–Minty theorem for variational inequalities.

Remark. Assumption (iv) in theorem 7 can be replaced by the following assumption:

(iv)* There exists a nonempty compact convex set B in K such that for every $x \in K \setminus B$, there exists $a \in B$ with

$$g(x,a) + h(x,a) \in -\operatorname{int} P. \tag{18}$$

Theorem 11. *Let the assumptions* (i)–(iii) *of theorem* 7 *and* (iv)* *hold. Then there exists* $x \in B$ *such that*

$$g(x,y) + h(x,y) \notin -\operatorname{int} P, \quad \forall y \in K.$$

Proof. Let $\{y_i : i \in N\}$ be a finite subset of K. Let

$$C := \operatorname{conv}\{B, \cup_{i \in N} y_i\}.$$

C is convex and compact. Hence, by lemmas 8 and 9, it follows that there exists at least one $x \in C$ such that

$$h(x,y) + g(x,y) \notin -\operatorname{int} P, \quad \forall y \in C. \tag{19}$$

For choosing $y := a \in B$, we obtain from (19) and (18) that $x \in B$. The P-monotonicity of g follows

$$h(x,y) - g(y,x) \notin -\operatorname{int} P, \quad \forall y \in C.$$

In particular

$$h(x,y_i) - g(y_i,x) \notin -\operatorname{int} P, \quad \forall i \in N.$$

As in the proof of lemma 8, it can be easily seen that every finite subfamily of the famil of closed sets

$$S(y) := \{x \in B : h(x,y) - g(y,x) \notin -\operatorname{int} P\} \text{ for each fixed, } y \in K$$

has nonempty intersection, and since B is compact then

$$\cap_{y \in K} S(y) \neq \phi.$$

Hence, there exists atleast one $x \in B$ such that

$$h(x,y) - g(y,x) \notin -\operatorname{int} P, \quad \forall y \in K.$$

From lemma 9, it follows that

$$h(x,y) - g(y,x) \notin -\operatorname{int} P, \quad \forall y \in K.$$

Finally we extend the notion of maximal monotonicity of a scalar multivalued mapping a vector multivalued mapping.

DEFINITION 12

A P-monotone multivalued mapping $T : K \longrightarrow 2^{L(X,Y)}$ is called P-maximal monotone, i for every pair

$$(u,x) \in L(X,Y) \times K : \langle v-u,\, y-x\rangle \notin -\operatorname{int} P, \quad \forall y \in K, \quad v \in Ty \Longrightarrow u \in$$

(2

By analogy of definition 12, we can define the following:

DEFINITION 13

A mapping $g : K \times K \longrightarrow Y$ with $g(x,x) = 0$, $\forall x \in K$ is called P-maximal monotone in t wide sense iff, for every pair

$$\begin{aligned}(u,x) \in L(X,Y) \times K : \langle -u, y-x\rangle - g(y,x) \notin -\operatorname{int} P, \quad \forall y \in K\\ \Longrightarrow g(x,y) - \langle u, y-x\rangle \notin -\operatorname{int} P, \quad \forall y \in K.\end{aligned} \qquad (2$$

Remark. If $g(x,y) := \sup_{u \in Tx}\langle u, y-x\rangle$ and T is P-maximal monotone, then g would P-maximal monotone according to this definition. However, to simplify matter, we add the following definition.

DEFINITION 14

A mapping $g : K \times K \longrightarrow Y$ with $g(x,x) = 0$, $\forall x \in K$ is called P-maximal monotone i for every $x \in K$ and for every P-convex mapping $\phi : K \longrightarrow Y$ with $\phi(x) = 0$:

$$\phi(y) - g(y,x) \notin -\operatorname{int} P, \quad \forall y \in K \Longrightarrow g(x,y) + \phi(y) \notin -\operatorname{int} P, \quad \forall y \in K.$$

(2

The relation of definitions 13 and 14 is the following:

Lemma 15. *Let $g : K \times K \longrightarrow Y$ be P-monotone; P-convex and lower semicontinuous the second argument, and linear Gateaux differentiable. Then definitions* 13 *and* 14 *b are equivalent.*

Proof. (21) $\Longrightarrow$ (22). Let $x \in K$, let $\phi : K \longrightarrow Y$ be P-convex with $\phi(x) = 0$, then

$$-\phi(y) - \langle \phi'(y), x - y \rangle \in P, \quad \forall y \in K. \tag{23}$$

where $\phi'(y)$ denotes linear Gateaux derivative of ϕ at y.

Assume that

$$\phi(y) - g(y, x) \notin -\operatorname{int} P, \quad \forall y \in K. \tag{24}$$

From (23), (24) and from (iv) of lemma 6, it follows that

$$-g(y, x) - \langle \phi'(y), x - y \rangle \notin -\operatorname{int} P, \quad \forall y \in K.$$

Let $y \in K$ be fixed, and set $x_\lambda := \lambda y + (1 - \lambda)x$, $\lambda \in (0, 1]$. Then $x_\lambda \in K$, and

$$-(1 - \lambda)g(x_\lambda, x) - (1 - \lambda)\langle \phi'(x_\lambda), x - x_\lambda \rangle \notin -\operatorname{int} P. \tag{25}$$

Hence

$$\begin{aligned} 0 = g(x_\lambda, x_\lambda) &\leq_P \lambda g(x_\lambda, y) + (1 - \lambda)g(x_\lambda, x) \\ &\Longrightarrow -(1 - \lambda)g(x_\lambda, x) \leq_P \lambda g(x_\lambda, x) \\ &\Longrightarrow -(1 - \lambda)g(x_\lambda, x) - (1 - \lambda)\langle \phi'(x_\lambda), x - x_\lambda \rangle \leq_P \lambda g(x_\lambda, y) \\ &\quad -(1 - \lambda)\langle \phi'(x_\lambda), x - x_\lambda \rangle. \end{aligned} \tag{26}$$

From (25), (26) and (iv) of lemma 6, we have

$$\lambda g(x_\lambda, y) - (1 - \lambda)\langle \phi'(x_\lambda), x - x_\lambda \rangle \notin -\operatorname{int} P.$$

Dividing by λ and using the P-monotonicity of g and (iv) of lemma 6, we get

$$-g(y, x_\lambda) + (1 - \lambda)\langle \phi'(x_\lambda), y - x \rangle \notin -\operatorname{int} P.$$

Letting $\lambda \longrightarrow 0_+$ and using the lower semicontinuity of g in the second argument, we obtain

$$-g(y, x) + \langle \phi'(x), y - x \rangle \notin -\operatorname{int} P, \quad \forall y \in K.$$

Since g is P-maximal monotone in the wide sense, we obtain

$$g(x, y) + \langle \phi'(x), y - x \rangle \notin -\operatorname{int} P \quad \forall y \in K. \tag{27}$$

But P-convexity of ϕ gives

$$\langle \phi'(x), y - x \rangle + g(x, y) - (g(x, y) + \phi(y)) \in P. \tag{28}$$

From (27), (28) and (iv) of lemma 6, we have

$$g(x, y) + \phi(y) \notin -\operatorname{int} P, \quad \forall y \in K.$$

(22) $\Longrightarrow$ (21). It follows immediately.

Now, we have the following theorem.

Theorem 16. *Let assumptions* (i) *and* (iii) *of Theorem* 7 *hold, and assume that the following conditions are satisfied:*

(ii)* $g : K \times K \longrightarrow Y$ *has the following properties:* $g(x, x) = 0$, $\forall x \in K$; g *is P-monotone and P-maximal monotone (defined as* (22)); g *is convex and lower semicontinuous in the second argument.*

(iv)* *There exists a nonempty compact convex subset B of K, such that for every $x \in K \setminus B$, there exists $a \in B$ such that*

$$-g(a,x) + h(x,a) \in -\operatorname{int} P.$$

Then there exists atleast one $x \in B$ such that

$$g(x,y) + h(x,y) \notin -\operatorname{int} P, \quad \forall y \in K. \tag{29}$$

Proof. Let $\{y_i : i \in N\}$, be a finite subset of K. Let $C := \operatorname{conv}\{B, \cup_{i\in N} y_i\}$. Then C is convex and compact. By lemma 8, it follows that there exists atleast one $x \in C$ such that

$$h(x,y) - g(y,x) \notin -\operatorname{int} P, \quad \forall y \in C. \tag{30}$$

Choosing $y := a \in B$, then (29) and (30) implies that $x \in B$. The P-monotonicity of g gives

$$h(x,y_i) - g(y_i,x) \notin -\operatorname{int} P, \quad \forall i \in N.$$

As in the proof of lemma 8, we have that every finite subfamily of the family of closed sets

$$S(y) := \{x \in B : h(x,y) - g(y,x) \notin -\operatorname{int} P\} \quad \text{for each fixed} \quad y \in K,$$

has nonempty intersection, and since B is compact then

$$\cap_{y\in K} S(y) \neq \phi,$$

i.e. there exists atleast one $x \in B$ such that

$$h(x,y) - g(y,x) \notin -\operatorname{int} P, \quad \forall y \in K. \tag{31}$$

Since g is P-maximal monotone, (31) implies

$$g(x,y) + h(x,y) \notin -\operatorname{int} P, \quad \forall y \in K.$$

Remark. Some results of this paper are the generalization of the results of [B–R].

References

[B–R] Blum E and Oettli W, From optimization and variational inequalities to equilibrium problems, *Math. Stud.* **63** (1994) 123–145

[C] Chen G-Y, Existence of solutions for a vector variational inequality: An extension of the Hartmann–Stampacchia Theorem, *J. Optim. Theor. Appl.* **74** (1992) 445–456

[C–C1] Chen G-Y and Craven B D, A vector variational inequality and optimization over an efficient set, *ZOR-Methods and Models of Operations Research* **34** (1990) 1–12

[C–C2] Chen G-Y and Craven B D, Existence and continuity for vector optimization, *J. Optim. Theor. Appl.* **81** (1994) 459–468

[F1] Fan K, A generalization of Tychonoff's fixed point theorem, *Math. Ann.* **142** (1961) 305–310

[F2] Fan K, A minimax inequality and applications, in: Inequalities III (ed.) O Shisha (New York: Academic Press) (1972) pp. 102–113

[G] Giannessi F, Theorems of alternative, quadratic programs and complementarity problems, in: Variational Inequalities and Complementarity Problems (eds) R W Cottle, F Giannessi and J L Lions (New York: Wiley) (1980) pp. 151–186

[K1] Kazmi K R, Existence of solutions for vector optimization, *Appl. Math. Lett.* **9** (1996) 19–22

[K2] Kazmi K R, Some remarks on vector optimization problems, *J. Optim. Theory Appl.* **96** (1998) 133–138

[K3] Kazmi K R, Existence of solutions for vector saddle point problems, in: Vector Variational Inequalities and Vector Equilibria. Mathematical Theories (ed.) F Giannessi (Dordrecht, Boston, London: Kluwer Academic Publishers) (2000) pp. 267–275

[K–A] Kazmi K R and Ahmad K, Nonconvex mappings and vector variational-like inequalities, in: Industrial and Applied Mathematics (ed.) A H Siddiqi and K Ahmad (New Delhi: Narosa Publishing House) (1998) pp. 103–115

[W–J] Weir T and Jeyakumar V, A class of nonconvex functions and mathematical programming, *Bull. Austral. Math. Soc.* **38** (1988) 177–189

[Y] Yang X-Q, Vector complementarity and minimal element problems, *J. Optim. Theory Appl.* **77** (1993) 483–495

Proc. Indian Acad. Sci. (Math. Sci.), Vol. 110, No. 2, May 2000, pp. 225–232.
© Printed in India

Existence of solutions of nonlinear integrodifferential equations of Sobolev type with nonlocal condition in Banach spaces

K BALACHANDRAN and K UCHIYAMA*

Department of Mathematics, Bharathiar University, Coimbatore 641 046, India
*Department of Mathematics, Sophia University, Tokyo 120-8554, Japan

MS received 16 September 1999

Abstract. In this paper we prove the existence of mild and strong solutions of a nonlinear integrodifferential equation of Sobolev type with nonlocal condition. The results are obtained by using semigroup theory and the Schauder fixed point theorem.

Keywords. Integrodifferential equation; Sobolev type; nonlocal condition; uniformly continuous semigroup; Schauder's fixed point theorem.

1. Introduction

The problem of existence of solutions of evolution equations with nonlocal conditions in Banach spaces has been studied first by Byszewski [7]. In that paper he has established the existence and uniqueness of mild, strong and classical solutions of the following nonlocal Cauchy problem:

$$u'(t) + Au(t) = f(t, u(t)), \qquad t \in (t_0, t_0 + a], \tag{1}$$

$$u(t_0) + g(t_1, t_2, \ldots, t_p, u(.)) = u_0, \tag{2}$$

where $-A$ is the infinitesimal generator of a C_0 semigroup $T(t)$, on a Banach space X, $0 \leq t_0 < t_1 < t_2 < \cdots < t_p \leq t_0 + a$, $a > 0$, $u_0 \in X$ and $f : [t_0, t_0 + a] \times X \to X$, $g : [t_0, t_0 + a]^p \times X \to X$ are given functions. Subsequently several authors have investigated the same type of problem to different classes of abstract differential equations in Banach spaces [1–4, 8, 11, 13, 14]. Brill [6] and Showalter [16] established the existence of solutions of semilinear evolution equations of Sobolev type in Banach spaces. This type of equations arise in various applications such as in the flow of fluid through fissured rocks [5], thermodynamics [9] and shear in second order fluids [10, 17]. The purpose of this paper is to prove the existence of mild and strong solutions for an integrodifferential equation of Sobolev type with nonlocal condition of the form

$$(Bu(t))' + Au(t) = f(t, u(t)) + \int_0^t g(t, s, u(s))\mathrm{d}s, \quad t \in (0, a], \tag{3}$$

$$u(0) + \sum_{k=1}^{p} c_k u(t_k) = u_0, \tag{4}$$

where $0 \leq t_1 < t_2 < \cdots < t_p \leq a$, B and A are linear operators with domains contained in a Banach space X and ranges contained in a Banach space Y and the nonlinear operators $f : I \times X \to Y$ and $g : \Delta \times X \to Y$ are given. Here $I = [0, a]$ and $\Delta = \{(s, t)$: $0 \leq s \leq t \leq a\}$.

2. Preliminaries

In order to prove our main theorem we assume certain conditions on the operators A and B. Let X and Y be Banach spaces with norm $|\cdot|$ and $\|\cdot\|$ respectively. The operators $A : D(A) \subset X \to Y$ and $B : D(B) \subset X \to Y$ satisfy the following hypothesis:

(H_1) A and B are closed linear operators,
(H_2) $D(B) \subset D(A)$ and B is bijective,
(H_3) $B^{-1} : Y \to D(B)$ is compact.

From the above fact and the closed graph theorem imply the boundedness of the linear operator $AB^{-1} : Y \to Y$. Further $-AB^{-1}$ generates a uniformly continuous semigroup $T(t), t \geq 0$ and so $\max_{t\in I} \|T(t)\|$ is finite. We denote $M = \max_{t\in I} \|T(t)\|$, $R = \|B^{-1}\|$. Let $B_r = \{x \in X : |x| \leq r\}$ and $c = \sum_{k=1}^{p} |c_k|$.

In this paper we assume that there exists an operator E on $D(E) = X$ given by the formula

$$E = \left[I + \sum_{k=1}^{p} c_k B^{-1} T(t_k) B\right]^{-1}$$

and $Eu_0 \in D(B)$,

$$E \int_0^{t_k} B^{-1} T(t_k - s) \left[f(s, u(s)) + \int_0^s g(s, \tau, u(\tau)) \mathrm{d}\tau\right] \mathrm{d}s \in D(B)$$

for $k = 1, 2, \ldots, p$.

The existence of E can be observed from the following fact (see also page 426 of [8]). Suppose that $\{T(t)\}$ is a C_0 semigroup of operators on X such that $\|B^{-1}T(t_k)B\| \leq Ce^{-\delta t_k} (k = 1, 2, \ldots, p)$, where δ is a positive constant and $C \geq 1$. If $\sum_{k=1}^{p} |c_k| e^{-\delta t_k} < 1/C$ then $\|\sum_{k=1}^{p} c_k B^{-1} T(t_k) B\| < 1$. So such an operator E exists on X.

DEFINITION 1 [15]

A continuous solution u of the integral equation

$$\begin{aligned} u(t) = {} & B^{-1}T(t)BEu_0 - \sum_{k=1}^{p} c_k B^{-1} T(t) BE \int_0^{t_k} B^{-1} T(t_k - s) \\ & \times \left[f(s, u(s)) + \int_0^s g(s, \tau, u(\tau)) \mathrm{d}\tau\right] \mathrm{d}s + \int_0^t B^{-1} T(t - s) \\ & \times \left[f(s, u(s)) + \int_0^s g(s, \tau, u(\tau)) \mathrm{d}\tau\right] \mathrm{d}s \end{aligned} \tag{5}$$

is said to be a mild solution of problem (3)–(4) on I.

DEFINITION 2 [15]

A function u is said to be a strong solution of problem (3)–(4) on I if u is differentiable almost everywhere on I, $u' \in L^1(I, X)$, $u(0) + \sum_{k=1}^{p} c_k u(t_k) = u_0$ and

$$(Bu(t))' + Au(t) = f(t, u(t)) + \int_0^t g(t, s, u(s))\mathrm{d}s, \quad \text{a.e on } I.$$

Remark. A mild solution of the nonlocal Cauchy problem (3)–(4) satisfies the condition (4). For, from (5)

$$u(0) = Eu_0 - \sum_{k=1}^{p} c_k E \int_0^{t_k} B^{-1}T(t_k - s)\left[f(s, u(s)) + \int_0^s g(s, \tau, u(\tau))\mathrm{d}\tau\right]\mathrm{d}s$$

and

$$\begin{aligned} u(t_i) = B^{-1}T(t_i)BEu_0 &- \sum_{k=1}^{p} c_k B^{-1}T(t_i)BE \int_0^{t_k} B^{-1}T(t_k - s) \\ &\times \left[f(s, u(s)) + \int_0^s g(s, \tau, u(\tau))\mathrm{d}\tau\right]\mathrm{d}s + \int_0^{t_i} B^{-1}T(t_i - s) \\ &\times \left[f(s, u(s)) + \int_0^s g(s, \tau, u(\tau))\mathrm{d}\tau\right]\mathrm{d}s. \end{aligned}$$

Therefore,

$$\begin{aligned} u(0) + \sum_{i=1}^{p} c_i u(t_i) &= \left[I + \sum_{i=1}^{p} c_i B^{-1}T(t_i)B\right] Eu_0 \\ &\quad - \left[I + \sum_{i=1}^{p} c_i B^{-1}T(t_i)B\right] \sum_{k=1}^{p} c_k E \int_0^{t_k} B^{-1}T(t_k - s) \\ &\quad \times \left[f(s, u(s)) + \int_0^s g(s, \tau, u(\tau))\mathrm{d}\tau\right]\mathrm{d}s + \sum_{i=1}^{p} c_i \\ &\quad \times \int_0^{t_i} B^{-1}T(t_i - s)\left[f(s, u(s)) + \int_0^s g(s, \tau, u(\tau))\mathrm{d}\tau\right]\mathrm{d}s \\ &= u_0 - \sum_{k=1}^{p} c_k \int_0^{t_k} B^{-1}T(t_k - s) \\ &\quad \times \left[f(s, u(s)) + \int_0^s g(s, \tau, u(\tau))\mathrm{d}\tau\right]\mathrm{d}s + \sum_{i=1}^{p} c_i \\ &\quad \times \int_0^{t_i} B^{-1}T(t_i - s)\left[f(s, u(s)) + \int_0^s g(s, \tau, u(\tau))\mathrm{d}\tau\right]\mathrm{d}s \\ &= u_0. \end{aligned}$$

Further assume that,

(H$_4$) $g : \Delta \times B_r \rightarrow Y$ is continuous in t and there exists a constant $K > 0$ such that

$$\|g(t, s, u)\| \leq K \quad \text{for} \quad (s, t) \in \Delta \quad \text{and} \quad u \in B_r.$$

(H$_5$) $f : I \times B_r \rightarrow Y$ is continuous in t on I and there exists a constant $L > 0$ such that

$$\|f(t, u)\| \leq L \quad \text{for} \quad t \in I \quad \text{and} \quad u \in B_r.$$

(H$_6$) $RM\|BEu_0\| + (R^2M^2a\|BE\|c + RMa)(L + Ka) \leq r.$

3. Main results

Theorem 1. *If the assumptions* $(\mathrm{H}_1) \sim (\mathrm{H}_6)$ *hold, then the problem* (3)–(4) *has a mild solution on I.*

Proof. Let $Z = C(I, X)$ and $Z_0 = \{u \in Z : u(t) \in B_r, t \in I\}$. Clearly, Z_0 is a bounded closed convex subset of Z. We define a mapping $F : Z_0 \to Z_0$ by

$$\begin{aligned}(Fu)(t) = {}& B^{-1}T(t)BEu_0 - \sum_{k=1}^{p} c_k B^{-1}T(t)BE \int_0^{t_k} B^{-1}T(t_k - s)\\ &\times \left[f(s, u(s)) + \int_0^s g(s, \tau, u(\tau))\mathrm{d}\tau\right]\mathrm{d}s + \int_0^t B^{-1}T(t-s)\\ &\times \left[f(s, u(s)) + \int_0^s g(s, \tau, u(\tau))\mathrm{d}\tau\right]\mathrm{d}s, \quad t \in I.\end{aligned}$$

Obviously F is continuous and maps Z_0 into itself. Moreover, F maps Z_0 into a precompact subset of Z_0. Note that the set $Z_0(t) = \{(Fu)(t) : u \in Z_0\}$ is precompact in X, for every fixed $t \in I$. We shall show that $F(Z_0) = S = \{Fu : u \in Z_0\}$ is an equicontinuous family of functions. For $0 < s < t$, we have

$$\begin{aligned}&\|(Fu)(t) - (Fu)(s)\| \leq \|B^{-1}(T(t) - T(s))BEu_0\|\\ &+ R^2Ma\|BE\|(L + Ka) \sum_{k=1}^{p} |c_k| \|(T(t) - T(s))\|\\ &+ \int_0^t \|B^{-1}\| \|T(t-\theta) - T(s-\theta)\| \left[\|f(\theta, u(\theta))\| + \int_0^\theta \|g(\theta, \tau, u(\tau))\|\mathrm{d}\tau\right]\mathrm{d}\theta\\ &+ \int_s^t \|B^{-1}\| \|T(s-\theta)\| \left[\|f(\theta, u(\theta))\| + \int_0^\theta \|g(\theta, \tau, u(\tau))\|\mathrm{d}\tau\right]\mathrm{d}\theta\\ &\leq (R\|BEu_0\| + R^2Ma\|BE\|(L + Ka)c)\|T(t) - T(s)\|\\ &+ R(L + Ka) \int_0^t \|T(t-\theta) - T(s-\theta)\|\mathrm{d}\theta + RM(L + Ka)|t - s|.\end{aligned}$$

The right hand side of the above inequality is independent of $u \in Z_0$ and tends to zero as $s \to t$ as a consequence of the continuity of $T(t)$ in the uniform operator topology for $t > 0$. It is also clear that S is bounded in Z. Thus by Arzela–Ascoli's theorem, S is precompact. Hence by the Schauder fixed point theorem, F has a fixed point in Z_0 and any fixed point of F is a mild solution of (3)–(4) on I such that $u(t) \in X$ for $t \in I$.

Next we prove that the problem (3)–(4) has a strong solution.

Theorem 2. *Assume that*

(i) *conditions* $(\mathrm{H}_1) \sim (\mathrm{H}_6)$ *hold*
(ii) *Y is a reflexive Banach space with norm* $\|\cdot\|$
(iii) $f : I \times B_r \to Y$ *is Lipschitz continuous in t, that is, there exists a constant* $L_1 > 0$ *such that*

$$\|f(t, u) - f(s, v)\| \leq L_1[|t - s| + \|u - v\|], \quad \textit{for } s, t \in I, \quad u, v \in B_r.$$

(iv) $g : \Delta \times B_r \to Y$ *is Lipschitz continuous in t, that is, there exists a constant* $L_2 > 0$

such that

$$\|g(t,\tau,u)-g(s,\tau,u)\| \le L_2|t-s|, \quad \textit{for } (t,\tau),(s,\tau)\in\Delta, \quad u\in B_r.$$

(v) *$Eu_0 \in D(AB^{-1})$ and*

$$E\int_0^{t_k} B^{-1}T(t_k-s)\left[f(s,u(s))+\int_0^s g(s,\tau,u(\tau))\mathrm{d}\tau\right]\mathrm{d}s \in D(AB^{-1})$$

$$\textit{for } k=1,\dots,p.$$

(vi) *u is the unique mild solution of the problem* (3)–(4).

Then u is a unique strong solution of problem (3)–(4) *on I.*

Proof. Since all the assumptions of Theorem 1 are satisfied, then the problem (3)–(4) has a mild solution belonging to $C(I,B_r)$. By assumption (vi), u is the unique mild solution of the problem (3)–(4). Now, we shall show that u is a unique strong solution of problem (3)–(4) on I.

For any $t\in I$, we have

$$\begin{aligned}
u(t+h)-u(t) &= B^{-1}[T(t+h)-T(t)]BEu_0 - \sum_{k=1}^p c_kB^{-1}[T(t+h)-T(t)]BE\\
&\quad\times\int_0^{t_k} B^{-1}T(t_k-s)\left[f(s,u(s))+\int_0^s g(s,\tau,u(\tau))\mathrm{d}\tau\right]\mathrm{d}s\\
&\quad+\int_0^{h} B^{-1}T(t+h-s)\left[f(s,u(s))+\int_0^s g(s,\tau,u(\tau))\mathrm{d}\tau\right]\mathrm{d}s\\
&\quad+\int_h^{t+h} B^{-1}T(t+h-s)\left[f(s,u(s))+\int_0^s g(s,\tau,u(\tau))\mathrm{d}\tau\right]\mathrm{d}s\\
&\quad-\int_0^{t} B^{-1}T(t-s)\left[f(s,u(s))+\int_0^s g(s,\tau,u(\tau))\mathrm{d}\tau\right]\mathrm{d}s\\
&= B^{-1}T(t)[T(h)-I]BEu_0 - \sum_{k=1}^p c_kB^{-1}[T(t+h)-T(t)]BE\\
&\quad\times\int_0^{t_k} B^{-1}T(t_k-s)\left[f(s,u(s))+\int_0^s g(s,\tau,u(\tau))\mathrm{d}\tau\right]\mathrm{d}s\\
&\quad+\int_0^{h} B^{-1}T(t+h-s)\left[f(s,u(s))+\int_0^s g(s,\tau,u(\tau))\mathrm{d}\tau\right]\mathrm{d}s\\
&\quad+\int_0^{t} B^{-1}T(t-s)[f(s+h,u(s+h))-f(s,u(s))]\mathrm{d}s\\
&\quad+\int_0^{t} B^{-1}T(t-s)\left[\int_0^{s+h} g(s+h,\tau,u(\tau))\mathrm{d}\tau-\int_0^s g(s,\tau,u(\tau))\mathrm{d}\tau\right]\mathrm{d}s.
\end{aligned}$$

Using our assumptions we observe that

$$\begin{aligned}
\|u(t+h)-u(t)\| &\le R\|BEu_0\|Mh\|AB^{-1}\| + cM^2R^2a\|BE\|(L+Ka)h\|AB^{-1}\|\\
&\quad+hRM(L+Ka)+RM\int_0^t L_1[h+\|u(s+h)-u(s)\|]\mathrm{d}s
\end{aligned}$$

$$+ RM\int_0^t\left[\int_0^s \|g(s+h,\tau,u)-g(s,\tau,u)\|\mathrm{d}\tau+\int_s^{s+h}\|g(s+h,\tau,u)\|\mathrm{d}\tau\right]\mathrm{d}s$$
$$\leq R\|BEu_0\|Mh\|AB^{-1}\|+[cM^2R^2a\|BE\|h\|AB^{-1}\|+hRM](L+Ka)$$
$$+RML_1\int_0^t[h+\|u(s+h)-u(s)\|]\mathrm{d}s+RMah(K+L_2a)$$
$$\leq Ph+Q\int_0^t\|u(s+h)-u(s)\|\mathrm{d}s,$$

where

$$P=R\|BEu_0\|M\|AB^{-1}\|+cM^2R^2a\|BE\|(L+Ka)\|AB^{-1}\|+RM(L+Ka)$$
$$+MRL_1a+RMKa+RML_2a^2$$

and $Q=RML_1$. By Gronwall's inequality

$$\|u(t+h)-u(t)\|\leq Phe^Q,\quad \text{for}\quad t\in J$$

Therefore, u is Lipschitz continuous on I. The Lipschitz continuity of u on I combinec with (iii) and (iv) imply that

$$t\longrightarrow f(t,u(t))\quad\text{and}\quad t\longrightarrow\int_0^t g(t,s,u(s))\mathrm{d}s$$

are Lipschitz continuous on I. Using the Corollary 2.11 in §4.2 of [15] and the definitior of strong solution we observe that the linear Cauchy problem:

$$(Bv(t))'+Av(t)=f(t,u(t))+\int_0^t g(t,s,u(s))\mathrm{d}s,\quad t\in(0,a]$$
$$v(0)=u_0-\sum_{k=1}^p c_ku(t_k)$$

has a unique strong solution v satisfying the equation

$$v(t)=B^{-1}T(t)Bv(0)+\int_0^t B^{-1}T(t-s)$$
$$\times\left[f(s,u(s))+\int_0^s g(s,\tau,u(\tau))\mathrm{d}\tau\right]\mathrm{d}s,\quad t\in I. \qquad (6$$

Now we will show that $v(t)=u(t)$ for $t\in I$. Observe that

$$v(0)=u(0)=Eu_0-\sum_{k=1}^p c_kE\int_0^{t_k}B^{-1}T(t_k-s)$$
$$\times\left[f(s,u(s))+\int_0^s g(s,\tau,u(\tau))\mathrm{d}\tau\right]\mathrm{d}s.$$

So,

$$B^{-1}T(t)Bv(0)=B^{-1}T(t)BEu_0-\sum_{k=1}^p c_kB^{-1}T(t)BE\int_0^{t_k}B^{-1}T(t_k-s)$$
$$\times\left[f(s,u(s))+\int_0^s g(s,\tau,u(\tau))\mathrm{d}\tau\right]\mathrm{d}s.$$

Substituting this in the equation (6) we see that $v(t) = u(t)$. Consequently, u is a strong solution of the problem (3)–(4) on I.

4. Example

Consider the following differential equation

$$\frac{\partial}{\partial t}(z(t,x) - z_{xx}(t,x)) - z_{xx}(t,x) = \mu(t, z(t,x)) + \int_0^t \eta(t,s,z(t,x))\mathrm{d}s \qquad x \in [0,\pi], t \in J, \tag{7}$$

$$z(t,0) = z(t,\pi) = 0, \quad t \in J$$

$$z(0,x) + \sum_{k=1}^{p} z(t_k, x) = z_0(x), \quad 0 < t_1 < t_2 < \cdots < t_p \leq a, \quad x \in [0,\pi]. \tag{8}$$

Let us take $X = Y = L^2[0,\pi]$. Define the operators $A : D(A) \subset X \to Y, B : D(B) \subset X \to Y$ by

$$Az = -z_{xx},$$
$$Bz = z - z_{xx},$$

respectively, where each domain $D(A), D(B)$ is given by

$$\{z \in X : z, z_x \text{ are absolutely continuous}, z_{xx} \in X, z(0) = z(\pi) = 0\}.$$

Define the operators $f : J \times X \to Y, \quad g : \Delta \times X \to Y$ by

$$f(t,z)(x) = \mu(t, z(t,x)), \quad g(t,s,z)(x) = \eta(t,s,z(t,x))$$

and satisfy the conditions (H_4) and (H_5) on a bounded closed set $B_r \subset X$. Here r satisfies the condition (H_6). Then the above problem (7) can be formulated abstractly as

$$(Bz(t))' + Az(t) = f(t,z) + \int_0^t g(t,s,z(s))\mathrm{d}s \quad t \in J.$$

Also, A and B can be written as [12]

$$Az = \sum_{n=1}^{\infty} n^2 \langle z, z_n \rangle z_n, \quad z \in D(A),$$

$$Bz = \sum_{n=1}^{\infty} (1 + n^2) \langle z, z_n \rangle z_n, \quad z \in D(B),$$

where $z_n(x) = \sqrt{2/\pi} \sin nx$, $n = 1, 2, \ldots$, is the orthoganal set of eigenfunctions of A. Furthermore, for $z \in X$ we have

$$B^{-1}z = \sum_{n=1}^{\infty} \frac{1}{(1+n^2)} \langle z, z_n \rangle z_n,$$

$$-AB^{-1}z = \sum_{n=1}^{\infty} \frac{-n^2}{(1+n^2)} \langle z, z_n \rangle z_n,$$

$$T(t)z = \sum_{n=1}^{\infty} \mathrm{e}^{-n^2 t/(1+n^2)} \langle z, z_n \rangle z_n.$$

It is easy to see that $-AB^{-1}$ generates a strongly continuous semigroup $T(t)$ on Y and $T(t)$ is compact such that $\|T(t)\| \leq e^{-t}$ for each $t > 0$. For this $T(t)$, B, B^{-1} we assume that the operator E exists. So all the conditions of the above theorem are satisfied. Hence the equation (7) with nonlocal condition (8) has a mild solution.

References

[1] Balachandran K and Chandrasekaran M, Existence of solutions of nonlinear integrodifferential equations with nonlocal condition, *J. Appl. Math. Stoch. Anal.* **10** (1997) 279–288

[2] Balachandran K and Chandrasekaran M, Existence of solutions of a delay differential equation with nonlocal condition, *Indian J. Pure. Appl. Math.* **27** (1996) 443–449

[3] Balachandran K, Park D G and Kwun Y C, Nonlinear integrodifferential equations of Sobolev type with nonlocal conditions in Banach spaces, *Comm. Korean Math. Soc.* **14** (1999) 223–231

[4] Balachandran K and Ilamaran S, Existence and uniqueness of mild and strong solutions of a Volterra integrodifferential equation with nonlocal conditions, *Tamkang J. Math.* **28** (1997) 93–100

[5] Barenblatt G, Zheltov I and Kochina I, Basic concepts in the theory of seepage of homogeneous liquids in fissured rocks, *J. Appl. Math. Mech.* **24** (1960) 1286–1303

[6] Brill H, A semilinear Sobolev evolution equation in Banach space, *J. Diff. Eq.* **24** (1977) 412–425

[7] Byszewski L, Theorems about the existence and uniqueness of solutions of a semilinear evolution nonlocal Cauchy problem, *J. Math. Anal. Appl.* **162** (1991) 494–505

[8] Byszewski L, Applications of properties of the right hand sides of evolution equations to an investigation of nonlocal evolution problems, *Nonlinear Anal.* **33** (1998) 413–426

[9] Chen P J and Curtin M E, On a theory of heat conduction involving two temperatures, *Z. Angew. Math. Phys.* **19** (1968) 614–627

[10] Huilgol R, A second order fluid of the differential type, *Internat. J. Nonlinear Mech.* **3** (1968) 471–482

[11] Jackson D, Existence and uniqueness of solutions to semilinear nonlocal parabolic equations, *J. Math. Anal. Appl.* **172** (1993) 256–265

[12] Lightbourne J H III and Rankin S M III, A partial functional differential equation of Sobolev type, *J. Math. Anal. Appl.* **93** (1983) 328–337

[13] Lin Y and Liu J U, Semilinear integrodifferential equations with nonlocal Cauchy problem, *Nonlinear Anal. TMA.* **26** (1996) 1023–1033

[14] Ntouyas S K and Tsamatos P Ch, Global existence of semilinear evolution equation with nonlocal conditions, *J. Math. Anal. Appl.* **210** (1997) 679–687

[15] Pazy A, Semigroups of linear operators and applications to partial differential equations, (New York: Springer–Verlag) (1983)

[16] Showalter R E, Existence and representation theorem for a semilinear Sobolev equation in Banach space, *SIAM J. Math. Anal.* **3** (1972) 527–543

[17] Ting T W, Certain nonsteady flows of second order fluids, *Arch. Rational Mech. Anal.* **14** (1963) 1–26

Proc. Indian Acad. Sci. (Math. Sci.), Vol. 110, No. 3, August 2000, pp. 233–261.
© Printed in India

Poincaré polynomial of the moduli spaces of parabolic bundles

YOGISH I HOLLA

School of Mathematics, Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400 005, India
E-mail: yogi@math.tifr.res.in

MS received 6 December 1999

Abstract. In this paper we use Weil conjectures (Deligne's theorem) to calculate the Betti numbers of the moduli spaces of semi-stable parabolic bundles on a curve. The quasi parabolic analogue of the Siegel formula, together with the method of Harder–Narasimhan filtration gives us a recursive formula for the Poincaré polynomials of the moduli. We solve the recursive formula by the method of Zagier, to give the Poincaré polynomial in a closed form. We also give explicit tables of Betti numbers in small rank, and genera.

Keywords. Cohomology; parabolic vector bundles; moduli space; Betti numbers; Weil conjectures.

Summary of notation

X	=	a smooth projective geometrically irreducible curve over the finite field $\mathbb{F}_q$.
$\bar{X}$	=	the curve $X \otimes_{\mathbb{F}_q} \bar{\mathbb{F}}_q$, where $\bar{\mathbb{F}}_q$ is an algebraic closure of $\mathbb{F}_q$.
$Z_X(t)$	=	the zeta function of the curve X.
X_ν	=	$X \otimes_{\mathbb{F}_q} \mathbb{F}_{q^\nu}$, where $\mathbb{F}_{q^\nu} \subset \bar{\mathbb{F}}_q$ is a finite field with q^ν elements. For positive integers n and m and non-negative integers $r_1, \ldots, r_m$ with $r_1 + \cdots + r_m = n$,
$\mathrm{Flag}(n, m, r_j)$	=	the variety of all flags $k^n = F_1 \supset \cdots \supset F_m \supset F_{m+1} = 0$ of vector subspaces in k^n, with $\dim(F_j/F_{j+1}) = r_j$.
$\lvert J(\mathbb{F}_q)\rvert$	=	the number of $\mathbb{F}_q$-rational points of the Jacobian of X.
S	=	a finite set of k-rational points of X. (These are the parabolic vertices.)
m_P	=	a fixed positive integer defined for each $P \in S$. For $P \in S$, and $1 \leq i \leq m_P$,
α	=	(α_i^P) is the set of allowed weights. For $P \in S$, and $1 \leq i \leq m_P$,
R	=	(R_i^P), the quasi-parabolic data (or simply 'data').
$n(R)$	=	$\sum_{i=1}^{m_P} R_i^P$, the rank of the data R.
L	=	a sub-data of R and
$R - L$	=	the complementary sub-data defined by $(R - L)_i^P = R_i^P - L_i^P$.
$\mathcal{L}$	=	a line bundle on X.
E	=	a vector bundle with a parabolic structure with data R.
$J_R(\mathcal{L})$	=	the set of isomorphism classes of quasi-parabolic vector bundles with data R, and determinant $\mathcal{L}$.

$\alpha(R)$ = $\sum_P \sum_{i=1}^{m_P} R_i^P \alpha_i^P$, the parabolic contribution to the degree.

$\deg(E)$ = the ordinary degree of E.

$\mathrm{pardeg}(E)$ = $\deg(E) + \alpha(R)$, the parabolic degree of E.

$\mathrm{par}\mu(E)$ = $\mathrm{pardeg}(E)/\mathrm{rank}(E)$, the parabolic slope of E. For $P \in S$, $1 \leq i \leq m_P$, and $1 \leq k \leq r$,

I = $(I_{i,k}^P)$, the intersection type of Nitsure, which is a partition of R.

$\ell(I)$ = the length of the intersection type I. For $j \leq r$, we have

R_j^I = the sub-data defined by $(R_j^I)_i^P = I_{i,j}^P$,

$R_{\leq j}^I$ = the sub-data defined by $(R_{\leq j}^I)_i^P = \sum_{k \leq j} I_{i,k}^P$,

$R_{\geq j}^I$ = the sub-data defined by $(R_{\geq j}^I)_i^P = \sum_{k \geq j} I_{i,k}^P$.

$I_{\leq j}$ = the partition of $R_{\leq j}^I$ defined by $(I_{\leq j})_{i,k}^P = I_{i,k}^P$ where $k \leq j$.

$I_{\geq j}$ = the partition of $R_{\geq j}^I$ defined by I $(I_{\geq j})_{i,k}^P = I_{i,k}^P$ where $k \geq j$.

$\mathcal{M}_{R,\mathcal{L}}$ = the moduli space of parabolic semi-stable bundles with the data R and determinant $\mathcal{L}$.

$\mathcal{M}_{R,\mathcal{L}}^s$ = the open sub variety of $\mathcal{M}_{R,\mathcal{L}}$ corresponding to the parabolic stable bundles. For parabolic bundles E', E and E'' with data R', R and R'', we denote by

$[E]$ = (E, i, j), a parabolic extension of E'' by E'.

$\mathrm{ParExt}(E'', E')$ the set of equivalence classes of parabolic extensions of E'' by E'.

$\beta_R(\mathcal{L})$ = $\sum(1/|\mathrm{ParAut}(E)|)$ where summation is over all $E \in J_R(\mathcal{L})$ such that E is parabolic semistable.

$J_R(\mathcal{L}, I)$ = the set of isomorphism classes of parabolic bundles with weights α, of intersection type I, and determinant $\mathcal{L}$.

$\beta_R(\mathcal{L}, I)$ = $\sum(1/|\mathrm{ParAut}(E)|)$, where the summation is over all E in $J_R(\mathcal{L}, I)$. We also write

$\beta_R(d, I)$ = $\beta_R(\mathcal{L}, I)$, since $\beta_R(\mathcal{L}, I)$ depends on $\mathcal{L}$ only via its degree $d = \deg(\mathcal{L})$.

$\mathcal{F}_R$ = $\prod_{P \in S} \mathrm{Flag}(n(R), m_P, R_i^P)$

$f_R(q)$ = the number of $\mathbb{F}_q$-valued points of the variety $\mathcal{F}_R$.

$C(I; d_1, \ldots, d_r)$ = the integer defined by equation (3.8).

$\sigma_k(I)$ = $\sum_{P \in S} \sum_{i>t} \sum_{l<r-k+1} I_{i,r-k+1}^P I_{t,l}^P$.

$\sigma_R(I)$ = $\sum_k \sigma_k(I)$. For a vector bundle F

$\chi(F)$ = the Euler characteristic.

$\chi\begin{pmatrix} \nu_1 & \cdots & \nu_r \\ \delta_1 & \cdots & \delta_r \end{pmatrix}$ = the numerical function of Desale–Ramanan defined by the equation (3.20).

$\sum_\circ$ denotes the summation over all $(d_1, \ldots, d_r) \in \mathbb{Z}^r$ with $\sum_i d_i = d$ and satisfying equation (3.7).

$\tau_{n(R)}(q)$ = $\frac{q^{(n(R)^2-1)(g-1)}}{q-1} Z_X(q^{-2}) \cdots Z_X(q^{-n(R)})$.

$\tilde{f}_R(t)$ = the rational function corresponding to f_R given by the equation (3.30).

$\tilde{\tau}_{n(R)}(t)$ = the rational function corresponding to $\tau_{n(R)}$ given by the equation (3.31).

$Q_{R,d}(t)$ = $t^{n(R)^2(g-1)}(1+t^{-1})^{2g}\tilde{\beta}_R(d)$, this is the main function for the recursion.

$Q_R(t)$ = $t^{n(R)^2(g-1)}\tilde{f}_R(t)\tilde{\tau}_{n(R)}(t)$.

$P_{R,d}$ = the power series whose coefficients compute the Betti numbers of the moduli space of parabolic stable bundles with data R and degree d.

$N_R(I; d_1, \dots, d_r)$ = the integer given by the formula (3.38).

Y = a smooth projective variety over $\mathbb{F}_q$.

N_ν = the number of $\mathbb{F}_{q^\nu}$-rational points of Y. For $i = 1, \dots, 2g$, we have

ω_i = a fixed algebraic integer of norm $q^{1/2}$.

$h(u, v_1, \dots, v_{2g})$ = a rational function given by the equation (4.2)

$p(u, v_1, \dots, v_{2g})$ = the numerator occuring in the equation (4.2).

$(a_{J.j})$ = the coefficients occuring in (4.3).

J = the multi-index $J = (i_1, i_2, \dots, i_{2g})$,

$|J|$ = $\sum_{r=1}^{2g} i_r$, and

v^J = $v_1^{i_1} v_2^{i_2} \cdots v_{2g}^{i_{2g}}$.

N = the 'weighted degree' of $p(u, v_1, v_2, \dots, v_{2g})$.

$b_{J,j}$ = the coefficients of h defined in (4.8).

$f_{\geq 0}(u, v_1, \dots, v_{2g})$ = the function defined by the equation (4.13).

M_r = $f_{\geq 0}(q^r, \omega_1^r, \dots, \omega_{2g}^r)$.

$Z_1(t)$ = the formal power series defined in (4.14).

$Z_2(t)$ = the formal power series defined in (4.15).

$Z(t)$ = $Z_1(t)Z_2(t)$. For a meromorphic function h on a disc in $\mathbb{C}$, and $\alpha > 0$,

$\mu(h, \alpha)$ = the number of zeros minus the number of poles counted with multiplicities of h with norm α.

$P(T)$ = the polynomial defined by the equation (4.19).

$M'_R(I; d)$ = the integer given by the formula (5.4). For a real number λ,

$M_R(I; \lambda)$ = the integer given by the formula (5.5).

$Q^\lambda_{R,d}(t)$ = the rational function defined in (5.7).

$S^\lambda_{R,d}(t)$ = the rational function defined in (5.8).

$\sum_{\circ\lambda}$ denotes the summation over $(d_1, \dots, d_r) \in \mathbb{Z}^r$ such that $\sum_i d_i = d$ and the equation (5.9) holds.

$Q^{\lambda^-}_{R,d}$ = $Q^{\lambda-\epsilon}_{R,d}$ for ϵ small enough such that the function $Q^\lambda_{R,d}$ has no jumps in the interval $[\lambda - \epsilon, \lambda)$.

$S^{\lambda^-}_{R,d}$ = $S^{\lambda-\epsilon}_{R,d}$ for ϵ small enough such that the function $S^\lambda_{R,d}$ has no jumps in the interval $[\lambda - \epsilon, \lambda)$.

$\Delta Q^\lambda_{R,d}$ = $Q^\lambda_{R,d} - Q^{\lambda^-}_{R,d}$.

$\Delta S^\lambda_{R,d}$ = $S^\lambda_{R,d} - S^{\lambda^-}_{R,d}$.

$\delta_R(L)$ = the integer given by the equation (5.10).

$d(\lambda, L)$ = $n(L)\lambda - \alpha(L)$.

$g_R(I; d)$ = the rational function given by (5.17).

$\sigma'_R(I)$ = $\sum_{P\in S} \sum_{k>l, i<t} I^P_{i,k} I^P_{t,l}$.

$M_g(I; \lambda)$ = the integer given by the formula (5.25).

$P_R(t)$ = the rational function (polynomial) defined by the equation (5.26). For a data R with rank $n(R) = 2$,

T = the subset of S consisting of parabolic vertices where the parabolic filtration is non-trivial.

T_I = $\{P \in T | I^P_{1,1} = 0\}$.

χ_I = a characteristic function on T, defined by the equation (6.3).

ψ_I = $\sum_{P\in T} \chi_I(P)(\alpha^P_1 - \alpha^P_2)$.

a_I = 1 if $d + [\psi_I]$ is even, and
= 0 if $d + [\psi_I]$ is odd.

δ^P = $\alpha^P_1 - \alpha^P_2$.

1. Introduction

This paper uses the Riemann hypothesis of Weil (Deligne's theorem) to explicitly determine the Betti numbers of the moduli of semistable parabolic bundles on a curve (when parabolic semi-stability implies parabolic stability).

Vector bundles with parabolic structures were introduced by Seshadri, and their moduli was constructed by Mehta–Seshadri (see [S] for an account). Our approach to the calculation of the Betti numbers is an extension of the method used by Harder and Narasimhan [HN] in the case of ordinary vector bundles. Harder and Narasimhan use the result of Siegel that the Tamagawa number of SL_r over a function field of transcendence degree one over a finite field is 1. This result can be reformulated purely in terms of vector bundles to give the formula (eq. (2.16)), which was used by Desale and Ramanan [DR] in their refinement of the Harder–Narasimhan Betti number calculation. In place of the above formula, we use its quasi-parabolic analogue (see eq. (2.19)) proved by Nitsure [N2], to extend the calculation of Harder and Narasimhan, as refined by Desale and Ramanan, to the parabolic case.

This gives us a recursive formula to obtain Betti numbers. Such a recursive formula had been obtained earlier for genus ≥ 2 by Nitsure [N1] using the Yang–Mills method of Atiyah–Bott [AB], and this was extended to lower genus by Furuta and Steer [FS].

Finally following Zagier's [Z] method of solving such a recursion (in the case of ordinary vector bundles), we obtain an explicit formula for the Poincaré polynomials. We give sample tables in lower ranks and genera.

This paper is arranged as follows. In § 2, we recall certain basic facts about parabolic bundles for the convenience of the reader. The paper of Desale and Ramanan computes the Poincaré polynomial of the moduli space of stable bundles, starting with the formula of Siegel (2.16). In § 3, we have followed their general pattern with suitable changes needed to handle the parabolic case, with the Siegel formula replaced by its parabolic analogue (2.19). This gives us the theorem (3.36), which is our desired recursive formula for the Poincaré polynomial. Along the way, we need a certain substitution ($\omega_i \to -t^{-1}$, $q \to t^{-2}$) used by Harder and Narasimhan, who have sketched its justification. We give a detailed proof of why such a substitution works (in a somewhat more general context) in § 4. In § 5, we solve the recursive formula using Zagier's [Z] approach, to get the explicit form (5.23) of the Poincaré polynomial. In § 6, we give some sample computations of the Poincaré polynomials of these moduli spaces and check their dependence on the weights and the degree when the rank is low (2, 3 and 4). In the appendix (§7), we have given tables for the Betti numbers of these moduli spaces in rank 2, 3 and 4.

2. Basic definitions and notations

Zeta function of a curve

Let $\mathbb{F}_q$ be a finite field, and let $\bar{\mathbb{F}}_q$ be its algebraic closure. Let X be a smooth projective geometrically irreducible curve over $\mathbb{F}_q$, where geometric irreducibility means $\bar{X} = X \otimes_{\mathbb{F}_q} \bar{\mathbb{F}}_q$ is irreducible.

Given any integer $r > 0$, let $\mathbb{F}_{q^r} \subset \bar{\mathbb{F}}_q$ be the unique field extension of degree r over $\mathbb{F}_q$. Let $N_r = |X(\mathbb{F}_{q^r})|$ be the cardinality of the set of $\mathbb{F}_q$-rational points of X. Recall that the zeta function of X is defined by

$$Z_X(t) = \exp\left(\sum_{r>0} \frac{N_r t^r}{r}\right). \tag{2.1}$$

By the Weil conjectures it follows that the zeta function has the form

$$Z_X(t) = \frac{\prod_{i=1}^{2g}(1-\omega_i t)}{(1-t)(1-qt)}, \tag{2.2}$$

where ω_i's are algebraic integers of norm $q^{1/2}$, and g is the genus of the curve. For $\nu \geq 1$, let X_ν denote the curve $X_\nu = X \otimes_{\mathbb{F}_q} \mathbb{F}_{q^\nu}$. The following remark will be used later.

Remark 2.3. If the zeta function of X over $\mathbb{F}_q$ is as given in (2.2), then the zeta function for the curve X_ν over $\mathbb{F}_{q^\nu}$ has the form

$$Z_{X_\nu}(t) = \frac{\prod_{i=1}^{2g}(1-\omega_i^\nu t)}{(1-t)(1-q^\nu t)}. \tag{2.4}$$

Rational points on flag varieties

Now we recall the computation of the number of rational points of flag varieties. Let $k = \mathbb{F}_q$ as before, let n and m be positive integers, and let there be given non-negative integers $r_1, \ldots, r_m$ with $r_1 + \cdots + r_m = n$. We denote by $\mathrm{Flag}(n, m, (r_j))$ the variety of all flags $k^n = F_1 \supset \cdots \supset F_m \supset F_{m+1} = 0$ of vector subspaces in k^n, with $\dim(F_j/F_{j+1}) = r_j$.

PROPOSITION 2.5

The number of $\mathbb{F}_q$*-rational points of* $\mathrm{Flag}(n, m, (r_j))$ *is*

$$f(q, n, m, (r_j)) = \frac{\prod_{i=1}^{n}(q^i - 1)}{\prod_{\{j|r_j \neq 0\}} \prod_{l=1}^{r_j}(q^l - 1)}. \tag{2.6}$$

Proof. The number of rational points $g(r,p)$ on the Grassmanian $\mathrm{Grass}(r,p)$ of p-dimensional subspaces of k^r can be seen to be

$$g(r,p) = \frac{(q^r - 1)\cdots(q^r - q^{p-1})}{(q^p - 1)\cdots(q^p - q^{p-1})} \tag{2.7}$$

and the number $f(q, n, m, (r_j))$ clearly satisfies

$$f(q, n, m, (r_j)) = g(n, r_m) g(n - r_m, r_{m-1}) \cdots g(r_1 + r_2, r_2). \tag{2.8}$$

Simplifying yields the desired formula. □

Parabolic vector bundles

Let $S = \{P_1, \ldots, P_s\}$ be any closed subset of X whose points are k-rational. For each $P \in S$, let there be given a positive integer m_P.

We fix an indexed family of real numbers (α_i^P), where $P \in S$ and $i = 1, \ldots, m_P$, satisfying $0 \leq \alpha_1^P < \alpha_2^P \cdots < \alpha_{m_P}^P < 1$, which we denote simply by α. We fix the set S, the integers (m_P) and the family α in all that follows.

The parabolic weights at P of the parabolic bundles that we will consider in this paper are going to belong to the chosen set $\{\alpha_1^P, \alpha_2^P, \ldots, \alpha_{m_P}^P\}$. (Note that this property will be inherited by the sub-quotients of such parabolic bundles.) This allows us to formulate the definition of parabolic bundles in a somewhat different way from Seshadri, which is more suited for our inductive arguments. However, the difference is only superficial, and we explain later (remark (2.11) on the next page) the bijective correspondence between parabolic bundles in our sense, and parabolic bundles in Seshadri's sense which have weights in our given set.

A *quasi-parabolic data* R (or simply 'data' when the context is clear) is an indexed family of non-negative integers (R_i^P) for $P \in S$ and $1 \leq i \leq m_P$, satisfying the following condition: $\sum_{i=1}^{m_P} R_i^P$ is a positive integer independent of $P \in S$. We call $n(R) = \sum_{i=1}^{m_P} R_i^P$ as the *rank* of the quasi-parabolic data R.

Let L be another quasi-parabolic data. We say L is a *sub-data* of a given data R if $L_i^P \leq R_i^P$for all P and i, and $n(L) < n(R)$. We also define its *complementary sub-data* $R - L$ by $(R-L)_i^P = R_i^P - L_i^P$. A quasi-parabolic structure with data R on a vector bundle E on X, by definition, consists of a flag

$$E^P = E_1^P \supset E_2^P \cdots \supset E_{m_P}^P \supset E_{m_P+1}^P = 0 \tag{2.9}$$

of vector subspaces in the fiber E^P over each point P of S, such that $R_i^P = \dim(E_i^P/E_{i+1}^P)$ for $P \in S$ and $1 \leq i \leq m_P$.

A parabolic structure with data R on a vector bundle E is a quasi-parabolic structure on E with data R along with weights α_i^P for each P and i. We say R_i^P is the multiplicity of the weight α_i^P. To a data R, we associate the real number $\alpha(R)$ by

$$\alpha(R) = \sum_P \sum_{i=1}^{m_P} R_i^P \alpha_i^P. \tag{2.10}$$

Remark 2.11. We record here the minor changes in notations and conventions that we have made (compared to the original notation of Seshadri). In our definition, note that the data R has the property that $R_i^P \geq 0$ (and not > 0), hence if E is a parabolic bundle in our sense with data R then the inclusions occurring in the filtration (2.9) are not necessarily strict. We recover the definition of Seshadri by re-defining the weights $\bar{\alpha}$ inductively as follows

$$\begin{aligned} \bar{\alpha}_1^P &= \min_i\{\alpha_i^P | R_i^P \neq 0\}, \\ \bar{\alpha}_j^P &= \min_k\{\alpha_k^P | R_k^P \neq 0, \alpha_k^P - \bar{\alpha}_{j-1}^P > 0\}. \end{aligned} \tag{2.12}$$

From this it follows that each $\bar{\alpha}_j^P$ equals α_i^P for exactly one i, which allows us to define $\bar{R}_j^P = R_i^P$ for that particular i. Now it is clear that E is a parabolic bundle in the sense of Seshadri with weights $\bar{\alpha}$ and multiplicities $\bar{R}$, where the flags are defined by sub-spaces $\bar{E}_j^P = E_i^P$ for that i for which $\bar{\alpha}_j^P = \alpha_i^P$. Since we have fixed the weights α, we can recover the parabolic bundle in our sense from a given parabolic bundle with weights $\bar{\alpha}$ and multiplicities $\bar{R}$ in the sense of Seshadri when the set $\{\bar{\alpha}_i^P\}$ is a subset of $\{\alpha_i^P\}$ for each P, by simply assigning

$$\begin{aligned} R_i^P &= 0 \text{ if } \alpha_i^P \neq \bar{\alpha}_j^P \text{ for any } j \\ &= \bar{R}_j^P \text{ if } \alpha_i^P = \bar{\alpha}_j^P \text{ for some } j \end{aligned} \tag{2.13}$$

and defining the sub-spaces occurring in the flags inductively by

$$\begin{aligned} E_1^P &= E^P, \\ E_i^P &= E_{i-1}^P \text{ if } \alpha_i^P \neq \bar{\alpha}_j^P \text{ for any } j \\ &= \bar{E}_j^P \text{ if } \alpha_i^P = \bar{\alpha}_j^P \text{ for some } j. \end{aligned} \tag{2.14}$$

This sets up a bijective correspondence between parabolic bundles in our sense and in the sense of Seshadri. Also note that

$$\sum_P \sum_i R_i^P \alpha_i^P = \sum_P \sum_i \bar{R}_i^P \bar{\alpha}_i^P, \tag{2.15}$$

which will enable us to write the parabolic degree in terms of our modified definition. The advantage of our definition is that it is easier to handle the induced parabolic structures on the sub-bundles and the quotient bundles in what follows. Also the parabolic homomorphisms between two parabolic bundles E and E' with data R and R' respectively in our sense, having the same fixed family of weights (α_i^P), are just filtration preserving homomorphisms of the vector bundles.

Quasi-parabolic Siegel formula

For a positive integer n and for any line bundle $\mathcal{L}$ on X, let $J_n(\mathcal{L})$ denote the set of isomorphism classes of vector bundles E on X with $\operatorname{rank}(E) = n$ and determinant $\mathcal{L}$. Let $|\operatorname{Aut}(E)|$ denote the cardinality of the group of all automorphisms of E. Then the Siegel formula, asserts that

$$\sum_{E \in J_n(\mathcal{L})} \frac{1}{|\operatorname{Aut}(E)|} = \frac{q^{(n^2-1)(g-1)}}{q-1} Z_X(q^{-2}) \cdots Z_X(q^{-n}). \tag{2.16}$$

The above formula was given a proof purely in terms of vector bundles by Ghione and Letizia [G-L].

For a line bundle $\mathcal{L}$ on X, let $J_R(\mathcal{L})$ denote the set of all isomorphism classes of quasi-parabolic vector bundles with data R, and determinant $\mathcal{L}$. Let $f_R(q)$ (denoted by $f(q, R)$ in [N2]) be the number of $\mathbb{F}_q$-valued points of the variety $\mathcal{F}_R = \prod_{P \in S} \operatorname{Flag}(n(R), m_P, (R_i^P))$ where $\operatorname{Flag}(n(R), m_P, (R_i^P))$ is the flag variety determined by (R_i^P). Now by equation (2.6), we have

$$f_R(q) = \frac{\prod_{i=1}^{n(R)} (q^i - 1)^{|S|}}{\prod_{P \in S} \prod_{\{i | R_i^P \neq 0\}} \prod_l^{R_i^P} (q^l - 1)}. \tag{2.17}$$

Let $|\text{Par Aut(E)}|$ denote the cardinality of the set of quasi-parabolic isomorphisms of a quasi-parabolic bundle E. The Siegel formula has the following quasi-parabolic analogue, which was proved by Nitsure [N2].

Theorem 2.18. (*Quasi-parabolic Siegel formula*)

$$\sum_{E \in J_R(\mathcal{L})} \frac{1}{|\text{Par Aut(E)}|} = f_R(q) \frac{q^{(n(R)^2-1)(g-1)}}{q-1} Z_X(q^{-2}) \cdots Z_X(q^{-n(R)}). \tag{2.19}$$

For example, if S is empty or more generally if the quasi-parabolic structure at each point of S is trivial (that is, each flag consists only of the zero subspace and the whole space), then on one hand $\operatorname{Par Aut}(E) = \operatorname{Aut}(E)$, and on the other hand each flag variety is a point, and so $f_R(q) = 1$. Hence in this situation the above formula reduces to the original Siegel formula.

Parabolic degree and stability

Let E be a parabolic bundle over X with data R. Because of (2.15), we can define the parabolic degree of E and the parabolic slope of E as follows:

$$\mathrm{pardeg}(E) = \deg(E) + \alpha(R) \text{ and } \mathrm{par}\mu(E) = \mathrm{pardeg}(E)/\mathrm{rank}(E). \tag{2.20}$$

A parabolic bundle E on X is said to be parabolic stable (resp. parabolic semi-stable) if for every non-trivial proper sub-bundle F of E with induced parabolic structure, we have $\mathrm{par}\mu(F) < \mathrm{par}\,\mu(E)$(resp. $\leq$).

The equation (2.15) implies that the definitions of parabolic stable (resp. parabolic semi-stable) bundles are not altered by the change in the definition of the parabolic bundles we have made.

We say that the numerical data (d, R) satisfies the condition 'par semi-stable = par stable' if every parabolic semistable bundle with data R and degree d is automatically parabolic stable.

Remark 2.21. If the degree d and rank $n(R)$ are coprime and all the weights are assumed to be very small ($\alpha_i^P < 1/(n(R)^2|S|)$ for example) then, each parabolic semistable bundle is actually parabolic stable.

We now recall the following.

Lemma 2.22. *If E is a parabolic stable bundle, then every parabolic homomorphism of E into itself is a scalar endomorphism.*

Parabolic Harder–Narasimhan intersection types

Recall the following.

PROPOSITION 2.23

Any non-zero parabolic bundle E with the data R admits a unique filtration by sub-bundles

$$0 = G_0 \not\subseteq G_1 \not\subseteq \cdots \not\subseteq G_r = E \tag{2.24}$$

satisfying

(i) G_i/G_{i-1} *is parabolic semistable for* $i = 1, \ldots, r$.
(ii) $\mathrm{par}\mu(G_i/G_{i-1}) > \mathrm{par}\mu(G_{i+1}/G_i)$ *for* $i = 1, \ldots, r-1$.

Equivalently,

(1) G_i/G_{i-1} *is parabolic semistable for* $i = 1, \ldots, r$.
(2) *For any parabolic sub-bundle F of E containing* G_{i-1} *we have* $\mathrm{par}\,\mu(G_i/G_{i-1}) > \mathrm{par}\,\mu(F/G_{i-1}), i = 1, \ldots, r$.

The result is first proved over an algebraically closed field, and then Galois descent is applied (using the uniqueness of the filtration) to prove that the filtration is defined over the original field.

The unique filtration is called the parabolic Harder–Narasimhan filtration. Let $(I_{i,k}^P)$ be an indexed collection of non-negative integers, where $P \in S$, $1 \leq i \leq m_P$, and $1 \leq k \leq r$

where r is a given positive integer. We say that $I = (I^P_{i,k})$ is a *partition* of R of length r if the following holds:

(1) For $P \in S$ and $1 \leq i \leq m_P$, we have $\sum_{k=1}^{r} I^P_{i,k} = R^P_i$.
(2) For $P \in S$ and $1 \leq k \leq r$, the summation $\sum_{i=0}^{m_P} I^P_{i,k}$ is independent of P.[1]
(3) Given any $k \leq r$, $I^P_{i,k} \neq 0$ for some P and i.

We write $\ell(I) = r$ to indicate that I has length r.

Suppose I is a partition of R with $\ell(I) = r$. For $j = 1, \dots, r$ define a sub-data R^I_j of R by the equality $(R^I_j)^P_i = I^P_{i,j}$. We also define the sub-data $R^I_{\leq j}$ (resp. $R^I_{\geq j}$) of R by the equality

$$(R^I_{\leq j})^P_i = \sum_{k \leq j} I^P_{i,k} \left(\text{resp. } (R^I_{\geq j})^P_i = \sum_{k \geq j} I^P_{i,k} \right). \tag{2.25}$$

Note that the rank $n(R^I_{\leq j})$ of $R^I_{\leq j}$ is equal to $n(R^I_1) + n(R^I_2) + \cdots + n(R^I_j)$. We observe that the partition I of R induces a partition $I_{\leq j}$ (resp. $I_{\geq j}$) on $R^I_{\leq j}$ (resp. $R^I_{\geq j}$) defined by $(I_{\leq j})^P_{i,k} = I^P_{i,k}$ (resp. $(I_{\geq j})^P_{i,k} = I^P_{i,k}$), where $k \leq j$ (resp. $k \geq j$).

We now recall how partitions, as abstractly defined above, are associated with parabolic bundles in Nitsure [N1]. To each $E \in J_R(\mathcal{L})$ we have the parabolic Harder–Narasimhan filtration $0 \subset G_1 \subset \cdots \subset G_r = E$ which gives a filtration on the fibers. Then the *intersection matrix* $(I^P_{i,k})$ corresponding to it is defined in [N1], by putting

$$I^P_{m_P,1} = \dim(E^P_{m_P} \cap G^P_1) \tag{2.26}$$

and

$$I^P_{j,l} = \dim(E^P_j \cap G^P_l) - \sum_{\substack{i \leq j, \text{ and } k \leq l \\ (i,k) \neq (l,j)}} I^P_{i,k}. \tag{2.27}$$

With this definition I becomes a partition of R with $\ell(I) = r$. The sub-bundles G_j (resp. quotients E/G_j) under the induced parabolic structure have the sub-data $R^I_{\leq j}$ (resp. $R^I_{\geq j+1}$). Also the sub-quotient G_j/G_{j-1} has the sub-data R^I_j.

Moduli spaces

For the moment assume that our ground field k is algebraically closed. Recall that parabolic semi-stable bundles, with a fixed parabolic slope, form an abelian category with the property that each object has finite length and simple objects are precisely the parabolic stable bundles. Hence for every parabolic semi-stable bundle E there exists a Jordan–Holder series

$$E = E_r \supset E_{r-1} \cdots \supset E_1 \supset 0$$

such that E_i/E_{i-1} is a parabolic stable bundle satisfying $\operatorname{par}\mu(E_i/E_{i-1}) = \operatorname{par}\mu(E)$. If we write $\operatorname{Gr}(E)$ for $\oplus_i E_i/E_{i-1}$, then it is well defined and is a parabolic semistable bundle with the same data as E. We say that two parabolic semi-stable bundles E and F are S-equivalent if $\operatorname{Gr}(E)$ and $\operatorname{Gr}(F)$ are isomorphic as parabolic bundles.

Mehta and Seshadri [M-S] prove that there exists a coarse moduli scheme $\mathcal{M}_{R,\mathcal{L}}$ of the S-equivalence classes of parabolic semistable bundles with the data R and determinant $\mathcal{L}$.

[1] For later reference, this number will be equal to the rank of G_k.

The scheme $\mathcal{M}_{R,\mathcal{L}}$ is a normal projective variety. Further the subset $\mathcal{M}^s_{R,\mathcal{L}}$ of $\mathcal{M}_{R,\mathcal{L}}$ corresponding to parabolic stable bundles is a smooth open subvariety.

Remark 2.28. Our method computes the Betti numbers of the moduli space of parabolic bundles for any curve over $\mathbb{C}$ because of the following reason. The theorem of Seshadri implies that the topological type of these moduli spaces depends only on the genus g, the cardinality of parabolic vertices $|S|$, the degree d and the set of weights α along with their multiplicities R. We start with such a data, construct a smooth projective absolutely irreducible curve X over a finite field $k=\mathbb{F}_q$ which has at least $|S|$ number of k-rational points (by taking $q=p^n$ for large n, or $q=p$ for a large prime p). The use of Witt vectors allows us to spread the curve and the moduli spaces to the quotient field of the ring of Witt vectors, when the condition 'par semistable = par stable' holds. Now by Weil conjectures it follows that the Betti numbers of the moduli space of parabolic bundles over X coincides with the one over the curve obtained by the change of base to $\mathbb{C}$

Parabolic extensions

Let E', E and E'' be parabolic bundles with data R', R and R'' respectively. Let

$$0 \longrightarrow E' \xrightarrow{i} E \xrightarrow{j} E'' \longrightarrow 0 \tag{2.29}$$

be a short exact sequence of the underlying vector bundles such that the parabolic structures induced on E' and E'' from the given parabolic structure on E coincide with the given parabolic structures of E' and E'', we say that (2.29) is a short exact sequence o parabolic bundles. We also say $[E]=(E,i,j)$ is a *parabolic extension* of E'' by E'.

We say two parabolic extensions $[E_1]$ and $[E_2]$ are equivalent if there exists an isomorphism of parabolic bundles $\gamma : E_1 \longrightarrow E_2$ such that the following diagram commutes:

$$\begin{tikzcd}
0 \arrow[r] & E' \arrow[r, "i_1"] \arrow[d, equal] & E_1 \arrow[r, "j_1"] \arrow[d, "\gamma"] & E'' \arrow[r] \arrow[d, equal] & 0 \\
0 \arrow[r] & E' \arrow[r, "i_2"] & E_2 \arrow[r, "j_2"] & E'' \arrow[r] & 0
\end{tikzcd} \tag{2.30}$$

We denote the set of equivalence classes of parabolic extensions by $\operatorname{Par Ext}(E'', E')$. Th proof of the following lemma is straight-forward and we omit it.

Lemma 2.31. *There is a canonical bijection between* $\operatorname{Par Ext}(E'', E')$ *and* $H^1(X, \operatorname{Par Hor}$ $(E'', E'))$, *where* $\operatorname{Par Hom}(E'', E'))$ is the sheaf of germs of parabolic homomorphism from E'' to E'.

By analogy with the case of ordinary vector bundles, we define an action of Par Au $(E'') \times \operatorname{Par Aut}(E')$ on $\operatorname{Par Ext}(E'', E')$ as follows: Given automorphisms $\alpha \in \operatorname{Par Aut}(E''$ $\beta \in \operatorname{Par Aut}(E')$ and a parabolic extension $[E]=(E,i,j) \in \operatorname{Par Ext}(E'', E')$ we define th parabolic extension $\beta[E]\alpha$ to be the extension $(E, \beta \circ i, j \circ \alpha)$.

Now fix a parabolic extension $[E]$ of E'' by E'. The proof of the following lemma i analogous to the corresponding statement for ordinary vector bundles.

Lemma 2.32. (a) *The orbit of* $[E]$ *under this action is the set of equivalence class o parabolic extensions which have their middle terms isomorphic to E as parabol bundles.*

(b) *The stablizer of* $[E]$ *under this action is precisely the subgroup of* $\operatorname{Par}\operatorname{Aut}(E'')\times \operatorname{Par}\operatorname{Aut}(E')$ *consisting of elements of the form* (α,β) *such that there exists a parabolic automorphism of* E *which takes* E' *to itself and induces* α *on* E'' *and* β *on* E'.

3. The inductive formula

The use of parabolic Harder–Narasimhan intersection types

In this section we use the quasi-parabolic Siegel formula to obtain a recursive formula for the Poincaré polynomial of the moduli space of parabolic stable bundles when the condition 'par semi-stable = par stable' holds.

The left hand side of the quasi-parabolic Siegel formula (2.19) can be split into the summations coming from the parabolic semi-stable bundles and the unstable ones. In view of this we first define

$$\beta_R(\mathcal{L}) = \sum \frac{1}{|\operatorname{Par}\operatorname{Aut}(E)|}, \tag{3.1}$$

where summation is over all $E \in J_R(\mathcal{L})$ such that E is parabolic semi-stable. We assume that the data R and degree d are so chosen that the condition 'par semi-stable = par stable' holds. In particular by lemma (2.22) this implies that for any such parabolic semi-stable bundle E, $|\operatorname{Par}\operatorname{Aut}(E)| = q-1$. Hence

$$|\mathcal{M}_{R,\mathcal{L}}(\mathbb{F}_q)| = (q-1)\beta_R(\mathcal{L}) \tag{3.2}$$

is the number of $\mathbb{F}_q$-rational points of the moduli space of parabolic semi-stable bundles with the data R and determinant $\mathcal{L}$.

Now we have to take care of the unstable part of the summation (2.19). This summation can be further split into parabolic Harder–Narasimhan intersection types. For these considerations, we make the following definitions:

Let I be a partition of R with $\ell(I) = r$. Let $J_R(\mathcal{L}, I)$ denote the set of isomorphism classes of parabolic bundles with data R, of intersection type I, and determinant $\mathcal{L}$.

Let

$$\beta_R(\mathcal{L}, I) = \sum \frac{1}{|\operatorname{Par}\operatorname{Aut}(E)|}, \tag{3.3}$$

where the summation is over all E in $J_R(\mathcal{L}, I)$. Note that $\beta_R(\mathcal{L}, I) = \beta_R(\mathcal{L})$ for the unique I which has $\ell(I) = 1$.

The summations occurring in (3.1) and (3.3) are finite because the parabolic bundles of fixed intersection type form a bounded family, so it is dominated by a variety, hence has only finitely many $\mathbb{F}_q$-rational points.

Now the quasi-parabolic Siegel formula (2.19) can be restated as

$$\sum_{r\geq 1}\sum_{\{I|\ell(I)=r\}} \beta_R(\mathcal{L}, I) = \frac{f_R(q)q^{(n(R)^2-1)(g-1)}}{q-1} Z_X(q^{-2})\cdots Z_X(q^{-n(R)}), \tag{3.4}$$

where $f_R(q)$ is given by (2.17).

Computation of the function $\beta_R(\mathcal{L}, I)$

The main step in the induction formula is to use the parabolic Harder–Narasimhan filtration to give a formula for $\beta_R(\mathcal{L}, I)$ when $\ell(I) > 1$, in terms of $\beta_{R'}(\mathcal{L}')$ of lower rank bundles. This we do in the following proposition which is an analogue of proposition (1.7) of Desale and Ramanan [D-R].

PROPOSITION 3.5

(a) *The numbers $\beta_R(\mathcal{L}, I)$ and $\beta_R(\mathcal{L})$ depend on $\mathcal{L}$ only via its degree $d = \deg(\mathcal{L})$ (hence they can be written as $\beta_R(d, I)$ and $\beta_R(d)$ resp.).*
(b) *$\beta_R(d, I)$ satisfies the following recursive relation*

$$\beta_R(d, I) = \sum_{\circ} q^{C(I;d_1,\ldots,d_r)} |J(\mathbb{F}_q)|^{r-1} \prod_{k=1}^{r} \beta_{R_k^I}(d_k), \tag{3.6}$$

where $\sum_\circ$ denotes the summation over all $(d_1, \ldots, d_r) \in \mathbb{Z}^r$ with $\sum_i d_i = d$ and satisfying the following inequalities

$$\frac{d_1 + \alpha(R_1^I)}{n(R_1^I)} > \frac{d_2 + \alpha(R_2^I)}{n(R_2^I)} > \cdots > \frac{d_r + \alpha(R_r^I)}{n(R_r^I)}. \tag{3.7}$$

Here $|J(\mathbb{F}_q)|$ denotes the number of $\mathbb{F}_q$-valued points of the Jacobian of X, and

$$C(I; d_1, \ldots, d_r) = \sum_{P \in S} \sum_{k>l; i>t} I_{i,k}^P I_{t,l}^P - \sum_{k>l} (d_l n(R_k^I) - d_k n(R_k^I)) + \sum_{k>l} n(R_l^I) n(R_k^I)(g-1). \tag{3.8}$$

Proof. We prove both parts ((a) and (b)) of the proposition simultaneously by induction on $n = n(R)$. If $\ell(I) = 1$ then there is nothing to prove.

Consider a parabolic bundle E with data R, admitting the parabolic Harder–Narasimhan filtration $0 \subset G_1 \subset \cdots \subset G_r = E$ of length $r \geq 2$. Let M be the quotient E/G_1. If we give the induced parabolic structure to M then it has the data $R_{\geq 2}^I$.

Let T be the set of equivalence classes of parabolic extensions of M by G_1 which has the property that the middle term is isomorphic to E as a parabolic bundle. By lemma (2.32 (a)) T is same as the orbit of $[E]$ under the action of $\operatorname{Par Aut}(M) \times \operatorname{Par Aut}(G_1)$ on $\operatorname{Par Ext}(M, G_1)$, hence

$$|T| = \frac{|\operatorname{Par Aut}(M)||\operatorname{Par Aut}(G_1)|}{|\text{stabilizer of}[E]|}. \tag{3.9}$$

Note that every parabolic automorphism of E takes G_1 to itself (hence also M). This implies that we get a group homomorphism

$$\operatorname{Par Aut}(E) \xrightarrow{\phi} \operatorname{Par Aut}(G_1) \times \operatorname{Par Aut}(M). \tag{3.10}$$

Now by lemma (2.32 (b)) the stabilizer of $[E]$ is the image of ϕ, while the kernel of ϕ is equal to $I_E + H^0(X, \mathcal{P}ar\,\mathcal{H}om(M, G_1))$. Combining all this we get

$$|T| = \frac{|\operatorname{Par Aut}(M)||\operatorname{Par Aut}(G_1)||\operatorname{Par Hom}(M, G_1)|}{|\operatorname{Par Aut}(E)|}. \tag{3.11}$$

By definition

$$\beta_R(\mathcal{L}, I) = \sum_{E \in J_R(\mathcal{L}, I)} \frac{1}{|\operatorname{Par Aut}(E)|} \tag{3.12}$$

which is

$$\sum_{(M,G_1)} \sum_{\mathscr{E}} \frac{1}{|\mathrm{Par\,Aut}(E)||T|}, \tag{3.13}$$

where the first summation extends over all pairs (M, G_1) with G_1, a parabolic semi-stable bundle with data R_1^I, and M, parabolic bundle with data $R_{\geq 2}^I$ and intersection type $I_{\geq 2}$, such that $\det(M) \otimes \det(G_1) = \det(E)$. The second summation extends over the set $\mathscr{E} = \mathrm{Par\,Ext}\,(M, G_1)$. By (3.11), the right hand side of the above expression (3.13) reduces to

$$\sum_{(M,G_1)} \frac{1}{|\mathrm{Par\,Aut}(M)||\mathrm{Par\,Aut}(G_1)|q^{\chi(\mathcal{P}\mathrm{ar}\mathcal{H}\mathrm{om}\,(M,G_1))}}, \tag{3.14}$$

where

$$\chi(\mathcal{P}\mathrm{ar}\mathcal{H}\mathrm{om}\,(M, G_1)) = \dim_{\mathbb{F}_q}(\mathrm{Par\,Hom}(M, G_1)) - \dim_{\mathbb{F}_q}(\mathrm{Par\,Ext}(M, G_1)) \tag{3.15}$$

is the Euler characteristic of the sheaf $\mathcal{P}ar\,\mathcal{H}om(M, G_1)$.

We define certain numerical functions which depend only on the partition I as follows:

$$\sigma_k(I) = \sum_{P \in S} \sum_{i>t} \sum_{l<r-k+1} I^P_{i,r-k+1} I^P_{t,l} \text{ and } \sigma_R(I) = \sum_k \sigma_k(I). \tag{3.16}$$

Then it can be checked that $\sigma_1(I)$ is the length of the torsion sheaf $\mathscr{S}_1(I)$, which is defined by the following exact sequence:

$$0 \longrightarrow \mathcal{P}ar\mathcal{H}om\,(M, G_1) \longrightarrow \mathcal{H}om\,(M, G_1) \longrightarrow \mathscr{S}_1(I) \longrightarrow 0. \tag{3.17}$$

Using the fact that

$$\chi(\mathcal{P}ar\mathcal{H}om\,(M, G_1)) = \chi(M^* \otimes G_1) - \sigma_1(I), \tag{3.18}$$

the sum (3.14) becomes

$$= \sum_{M,G_1} \frac{q^{\sigma_1(I)}}{|\mathrm{Par\,Aut}(M)||\mathrm{Par\,Aut}(G_1)|q^{\chi(M^* \otimes G_1)}}. \tag{3.19}$$

Recall that Desale–Ramanan [D–R] introduced certain numerical functions

$$\chi\begin{pmatrix} \nu_1 & \cdots & \nu_r \\ \delta_1 & \cdots & \delta_r \end{pmatrix} = \sum_{k>l} (\delta_l \nu_k - \delta_k \nu_k) + \sum_{k>l} \nu_l \nu_k (g-1). \tag{3.20}$$

With this definition of χ, we have the following equality

$$\chi\begin{pmatrix} n(R_1^I) & n(R) - n(R_1^I) \\ d_1 & d - d_1 \end{pmatrix} = \chi(M^* \otimes G_1) \tag{3.21}$$

as in [D–R]. Now by (3.21), the sum (3.19) equals

$$\sum_{d_1} q^{\sigma_1(I) - \chi\begin{pmatrix} n(R_1^I) & n(R) - n(R_1^I) \\ d_1 & d - d_1 \end{pmatrix}} \sum_{(\eta,\gamma)} \sum_M \frac{1}{\mathrm{Par\,Aut}(M)} \sum_{G_1} \frac{1}{\mathrm{Par\,Aut}(G_1)}, \tag{3.22}$$

where the first summation in (3.22) is over all integers d_1 with

$$(d_1 + \alpha(R^I_1))/n(R^I_1) > (d - d_1 + \alpha(R^I_{\geq 2}))/(n - n(R^I_1)). \tag{3.23}$$

The second summation in (3.22) is over isomorphism classes of line bundles η and γ such that $\eta \otimes \gamma = \mathcal{L}$. The third one is over all parabolic bundles M with data R, having intersection type $I_{\geq 2}$, and determinant η. The fourth summation in (3.22) is over all semistable parabolic bundles G_1 with data R^I_1, and determinant γ. This expression is equal to

$$\sum q^{\sigma_1(I) - \chi\begin{pmatrix} n(R^I_1) & n - n(R^I_1) \\ d_1 & d - d_1 \end{pmatrix}} \sum \beta_{R^I_{\geq 2}}(\eta, I_{\geq 2}) \beta_{R^I_1}(\gamma). \tag{3.24}$$

Now note that by induction, the terms inside the summation are independent of $\mathcal{L}$, hence part (a) of the proposition follows. From now on we write $\beta_R(d)$ and $\beta_R(d,I)$ for $\beta_R(\mathcal{L})$ and $\beta_R(\mathcal{L}, I)$.

By Desale–Ramanan [D–R] we have the relation

$$\chi\begin{pmatrix} n(R^I_1) & n - n(R^I_1) \\ d_1 & d - d_1 \end{pmatrix} + \chi\begin{pmatrix} n(R^I_2) & \cdots & n(R^I_r) \\ d_2 & \cdots & d_r \end{pmatrix} = \chi\begin{pmatrix} n(R^I_1) & \cdots & n(R^I_r) \\ d_1 & \cdots & d_r \end{pmatrix}. \tag{3.25}$$

Using this and the induction hypothesis for $\beta_{R^I_{\geq 2}}(d - d_1, I_{\geq 2})$ we obtain the following equality:

$$\beta_R(d,I) = \sum q^{\sigma_R(I) - \chi\begin{pmatrix} n(R^I_1) & \cdots & n(R^I_r) \\ d_1 & \cdots & d_r \end{pmatrix}} |J_{\mathbb{F}_q}|^{r-1} \prod_{k=1}^{r} \beta_{R^I_k}(d_k). \tag{3.26}$$

As

$$C(I; d_1, \ldots, d_r) = \sigma_R(I) - \chi\begin{pmatrix} n(R^I_1) & \cdots & n(R^I_r) \\ d_1 & \cdots & d_r \end{pmatrix} \tag{3.27}$$

the proof of the proposition is complete. □

The recursive formula

The inductive expression for $\beta_R(d)$ can now be written as

$$f_R(q)\tau_{n(R)}(q) - \sum_{r \geq 2} \sum_{\{I | \ell(I) = r\}} \sum_{\circ} q^{C(I; d_1, \ldots, d_r)} |J_{\mathbb{F}_q}|^{r-1} \prod_{k=1}^{r} \beta_{R^I_k}(d_k), \tag{3.28}$$

where

$$\tau_{n(R)}(q) = \frac{q^{(n(R)^2 - 1)(g-1)}}{q-1} Z_X(q^{-2}) \cdots Z_X(q^{-n(R)}). \tag{3.29}$$

We now base change from $\mathbb{F}_q$ to $\mathbb{F}_{q^\nu}$. For the curve X_ν defined in § 2, the $\beta_R(d, q^\nu)$ will be a function of q^ν and ω_i^ν for $i = 1, \ldots, 2g$.

It follows from induction formula and eq. (2.17) that the function $\beta_R(d, q^\nu)$ is a polynomial in ω_i^ν for $i = 1, \ldots, 2g$, and is a rational function in q^ν with the property that the denominator has factors only of the form $q^{\nu n_0}(q^{\nu n_1} - 1)(q^{\nu n_2} - 1) \cdots (q^{\nu n_k} - 1)$, with $n_i \geq 1$ for $i \geq 1$. For such a function, one can substitute $-t^{-1}$ for ω_i and t^{-2} for q, to obtain a new rational function. We denote this operation by $\phi \to \tilde{\phi}$. For example

$$\tilde{f}_R(t) = \frac{t^{-2\dim \mathcal{F}_R} \prod_{i=1}^{n(R)} (1-t^{2i})^{|S|}}{\prod_{P\in S} \prod_{\{i | R_i^P \neq 0\}} \prod_{l=1}^{R_i^P} (1-t^{2l})} \tag{3.30}$$

and $\tilde{\tau}_{n(R)}(t)$ can be computed to be

$$\frac{t^{-2n(R)^2(g-1)} \prod_{i=1}^{n(R)} (1+t^{2i-1})^{2g}}{(1-t^{2n(R)}) \prod_{i=1}^{n(R)-1} (1-t^{2i})^2}. \tag{3.31}$$

This substitution is an important step in the computation of the Poincaré polynomial for the moduli space because of the proposition (4.34) of the next section.

Now we shall define rational functions $Q_{R,d}(t)$ and $Q_R(t)$ by

$$Q_{R,d}(t) = t^{n(R)^2(g-1)}(1+t^{-1})^{2g}\tilde{\beta}_R(d) \tag{3.32}$$

and

$$Q_R(t) = t^{n(R)^2(g-1)}\tilde{f}_R(t)\tilde{\tau}_{n(R)}(t). \tag{3.33}$$

Observe that if the condition 'par semi-stable = par stable' is satisfied and if we define

$$P_{R,d}(t) = t^{2\dim \mathcal{F}_R + 2(n(R)^2-1)(g-1)}(t^{-2}-1)\tilde{\beta}_R(d), \tag{3.34}$$

then we have the relation

$$P_{R,d}(t) = \frac{t^{2\dim \mathcal{F} + n(R)^2(g-1)}(1-t^2)}{(1+t)^{2g}} Q_{R,d}(t). \tag{3.35}$$

Now by proposition (4.34) and the fact that the dimension of the moduli space of parabolic semi-stable bundles is equal to $\dim \mathcal{F}_R + (n(R)^2-1)(g-1)$, we get that $P_{R,d}$ is a power series in t which computes the Betti numbers of the moduli space of parabolic stable bundles with the given data R and degree d.

If we perform the tilde operation on the original formula, we get the following recursive formula.

Theorem 3.36. *The functions* $Q_{R,d}$ *and* Q_R *defined by* (3.32) *and* (3.33) *satisfy the following recursion formula*

$$Q_R(t) = \sum_{r\geq 1} \sum_{\{I | \ell(I)=r\}} \sum_{\circ} t^{2N_R(I;d_1,\ldots,d_r)} \prod_{k=1}^{r} Q_{R_k^I,d_k}(t), \tag{3.37}$$

where the second summation extends over all partitions I *of* R *of length* r, *and where*

$$N_R(I;d_1,\ldots,d_r) = \sum_{k>l} (d_l n(R_k^I) - d_k n(R_l^I)) - \sum_{P\in S} \sum_{k>l,i>t} I_{i,k}^P I_{t,l}^P. \tag{3.38}$$

4. The substitution $\omega_i \to -t^{-1}$, $q \to t^{-2}$

In this section, we justify the substitution $\omega_i \to -t^{-1}$ and $q \to t^{-2}$, which gives us a recipe to compute the Poincaré polynomial of the moduli spaces, directly from the computation of the $\mathbb{F}_q$-rational points.

This substitution was briefly sketched in [H–N] for the rational function which counted the $\mathbb{F}_q$ rational points of the moduli space of stable bundles when rank and degree are coprime. We formulate and prove this in a more general setup, which we have used in the body of the paper.

Let Y be a smooth projective variety over $\mathbb{F}_q$. Let $N_\nu = |Y(\mathbb{F}_{q^\nu})|$ and let $\omega_1, \ldots, \omega_{2g}$ be fixed algebraic integers of norm $q^{1/2}$. Our basic assumption is that N_ν is given by some formula

$$N_\nu = h(q^r, \omega_1^r, \ldots, \omega_{2g}^r), \tag{4.1}$$

where $h(u, v_1, \ldots, v_{2g})$ is a rational function of the form

$$\frac{p(u, v_1, \ldots, v_{2g})}{u^{n_0}(u^{n_1} - 1) \ldots (u^{n_k} - 1)}, \tag{4.2}$$

where $p(u, v_1, \ldots, v_{2g}) \in \mathbb{Z}[u, v_1, \ldots, v_{2g}]$ is a polynomial with integral coefficients, and where $n_i \geq 1$ for all $i > 0$ and $n_0 \geq 0$. We wish to write down the Poincaré polynomial of Y in terms of the function h.

We first write down the function h as a suitable series and bound the coefficients. We can expand the numerator occurring in the expression for h as

$$p(u, v_1, \ldots, v_{2g}) = \sum_{l=0}^{N} \sum_{|J|+2j=l} a_{J,j} v^J u^j, \tag{4.3}$$

where J denotes the multi-index $J = (i_1, \ldots, i_{2g})$, $|J| = \sum_{r=1}^{2g} i_r$, and $v^J = v_1^{i_1} \ldots v_{2g}^{i_{2g}}$. Let $C > 0$ be any fixed integer such that $|a_{J,j}| < C$ for all J,j. The integer N in the summation above can be taken to be the 'weighted degree' of $p(u, v_1, \ldots, v_{2g})$ where the variable u is given weight 2.

We can rewrite h as

$$\frac{1}{u^n(1 - u^{-n_1}) \cdots (1 - u^{-n_k})} \sum_{l=0}^{N} \sum_{|J|+2j=l} a_{J,j} v^J u^j, \tag{4.4}$$

where $n = \sum_{i=0}^{k} n_i$. Expanding each $1/(1 - u^{-n_i})$ as a power series in u_{-1}, we get

$$h = \frac{1}{u^n} \sum_{l=0}^{N} \sum_{|J|+2j'=l} \sum_{i \leq 0} a_{J,j'} b_i v^J u^{j'+i}, \tag{4.5}$$

where b_i is the cardinality of the set of k-tuples of non-negative integers $(a_1, \ldots, a_k)$ such that $\sum_{r=1}^{k} a_r n(R) = -i$. Clearly, we have

$$b_i \leq (-i + 1)^k. \tag{4.6}$$

Now the right hand side of the equation (4.5) becomes

$$\frac{1}{u^n} \sum_{l \leq N} \sum_{|J|+2j=l} \left(\sum_{j'+i=j} a_{J,j} b_i \right) v^J u^j. \tag{4.7}$$

Define

$$b_{J,j} = \sum_{j'+i=j} a_{J,j} b_i. \tag{4.8}$$

This is a finite sum, which makes sense for every J,j such that $|J|+2j \leq N$. In terms of these $b_{J,j}$, the expression for h can be written as

$$h = \frac{1}{u^n} \sum_{l \leq N} \sum_{|J|+2j=l} b_{J,j} v^J u^j. \tag{4.9}$$

Note that in the above series, there are only finitely many positive powers of u and infinitely many negative powers. The following lemma puts a bound on the coefficients $b_{J,j}$.

Lemma 4.10. *The coefficients $b_{J,j}$ as defined above satisfies the following inequality*

$$|b_{J,j}| \leq CN(N-j+1)^k. \tag{4.11}$$

Proof. One observes that

$$|b_{J,j}| \leq \sum_{j'+i=j} |a_{J,j} b_i| \leq C \sum_{j'+i=j} |b_i|, \tag{4.12}$$

where j' and i are as in the preceding discussion. In the last expression of (4.12), the number of terms is $\leq N$, and by (4.6) each term $|b_i|$ is bounded by $(-i+1)^k$. As $(-i+1)^k \leq (N-j+1)^k$ for $j'+i=j$, the last expression (4.12) is bounded by $CN(N-j+1)^k$. This proves the lemma. □

Let

$$\begin{aligned} h_{\geq 0}(u, v_1, \ldots, v_{2g}) &= \textstyle\sum_{l=2n}^{N} \sum_{|J|+2j=l} b_{J,j} v^J u^{j-n} \quad \text{if } N \geq 2n \\ &= 0 \qquad\qquad\qquad\qquad\qquad\qquad \text{otherwise} \end{aligned} \tag{4.13}$$

and M_r be $h_{\geq 0}(q^r, \omega_1^r, \omega_2^r, \ldots, \omega_{2g}^r)$, then these numbers are well defined because of lemma 1.

Let

$$Z_1(t) = \exp\left(\sum_{r \geq 1} M_r t^r / r\right) \tag{4.14}$$

and

$$Z_2(t) = \exp\left(\sum_{r \geq 1} (N_r - M_r) t^r / r\right), \tag{4.15}$$

then $Z_1(t)$ and $Z_2(t)$ are well defined formal power series. We also define

$$Z(t) = Z_1(t) Z_2(t). \tag{4.16}$$

Given any meromorphic function h on a disc in $\mathbb{C}$, let $\mu(h, \alpha)$ denote the number of zeros minus the number of poles h with norm α, counted with multiplicities.

Lemma 4.17. (a) *$Z_2(t)$ is a non-vanishing holomorphic function on the disc $|t| < q^{1/2}$ and $Z_1(t)$ is a rational function, hence $Z(t)$ is a well defined meromorphic function in the region $|t| < q^{1/2}$, such that*

$$\mu(Z(t), q^{-i/2}) = \mu(Z_1(t), q^{-i/2}). \tag{4.18}$$

(b) *Let*

$$P(T) = \sum_{i\geq 0}(-1)^{i+1}\mu(Z(t), q^{-i/2})T^i, \qquad (4.19$$

then

$$P(T) = h_{\geq 0}(T^2, -T, -T, \ldots, -T). \qquad (4.20$$

Proof. To prove that the function $Z_2(t)$ has the above mentioned property it is enough t verify that the function

$$g(t) := \sum_{r\geq 1}(N_r - M_r)t^r/r \qquad (4.21$$

is holomorphic on the disc $|t| < q^{1/2}$. This function is

$$\sum_{l<2n}\sum_{|J|+2j=l} b_{J,j}\omega^{Jr}q^{jr-nr}t^r/r. \qquad (4.22$$

The coefficient of t^r is equal to

$$\sum_{l<2n}\sum_{|J|+2j=l} b_{J,j}\omega^{Jr}q^{jr-nr}/r \qquad (4.23$$

whose modulus is bounded by

$$\sum_{l<2n}\sum_{|J|+2j=l} |b_{J,j}|q^{r(j-n+|J|/2)}/r. \qquad (4.24$$

This by lemma (4.1) is

$$\leq \frac{NC}{rq^{nr}}\sum_{l<2n}\sum_{|J|+2j=l}(N-j+1)^k q^{r(2j+|J|)/2)} \qquad (4.2$$

$$\leq \frac{N^2C}{rq^{nr}}\sum_{l<2n} q^{rl/2}((3N+2-l)/2)^k \qquad (4.2$$

$$\leq \frac{N^2C}{2^k r}\sum_{l>0} q^{-rl/2}(3N+2-2n+l)^k \qquad (4.2$$

which is clearly a finite sum for $r \geq 1$ because powers of q decay exponentially and t other term has polynomial growth. Now since $(a+l) \leq a^l$ for $a \geq 2$ therefore the abo summation is bounded by

$$2^{-k}N^2C\sum_{l>0}((3N+2-2n)^k/q^{r/2})^l/r. \qquad (4.2$$

Suppose r is large enough such that $q^{r/2} > 2(3N+2-2n)^k$ the coefficient of t^r has t bound $2^{-k+1}N^2C(3N+2-2n)^k/(rq^{r/2})$ and the series with coefficient of t^r as above f large r clearly has radius of convergence $q^{r/2}$. Now we compute $Z_1(t)$ as

$$\exp\left(\sum_{r\geq 1}\sum_{l=2n}^{N}\sum_{|J|+2j=l} b_{J,j}\omega^{Jr}q^{jr-nr}t^r/r\right) \qquad (4.2$$

$$=\prod_{l=2n}^{N}\prod_{|J|+2j=l}\exp\left(b_{J,j}\sum_{r\geq 1}\omega^{Jr}q^{jr-nr}t^r/r\right) \tag{4.30}$$

which is equal to

$$\prod_{2n\leq l\leq N}\prod_{|J|+2j=l}(1-\omega^J q^{j-n}t)^{(-1)b_{J,j}}, \tag{4.31}$$

hence this is a rational function, and this also proves that $Z(t)$ is a meromorphic function in the region $|t| < q^{1/2}$, and that $\mu(Z(t), q^{-i/2}) = \mu(Z_1(t), q^{-i/2})$. This finishes the proof of part (a).

Also from here we can read off that

$$\mu(Z_1(t), q^{-i/2}) = (-1)\sum_{|J|+2j+2n=i} b_{J,j}. \tag{4.32}$$

Clearly the polynomial $f_{\geq 0}(T^2, -T, -T, \ldots, -T)$ now coincides with

$$\sum_{i\geq 0}(-1)^{i+1}\mu(Z_1(t), q^{-i/2})T^i. \tag{4.33}$$

Now by part (a) the proof of the lemma is complete. □

If the $\mathbb{F}_q$-rational points of the variety Y are given by the eqs (4.1) and (4.2), then $Z(t)$ is the zeta function of Y and $P(t)$ is the Poincaré polynomial of Y. We can restate the lemma (4.17) in terms of the Poincaré polynomial of Y, using Poincaré duality, as follows.

PROPOSITION 4.34

The function $T^{2\dim(Y)}h(T^{-2}, -T^{-1}, -T^{-1}, \ldots, -T^{-1})$ has a formal power series expansion $\sum_{\nu\geq 0} b_\nu T^\nu$ where b_ν is the νth-Betti number of Y for $\nu \leq 2\dim(Y)$.

5. The closed formula

In this section we solve the recursion formula (theorem (3.36)) to obtain a closed formula for the Poincaré polynomial of the moduli space of parabolic stable bundles under the condition 'par semi-stable = par stable'. We do this by generalizing the method of Zagier [Z] to the parabolic set up.

The induction formula can be re-written as

$$Q_R(x) = \sum_{r\geq 1}\sum_{I}\sum_{\circ} x^{n(R)(I;d_1,\ldots,d_r)}\prod_{k=1}^{r} Q_{R_k^I,d_k}(x), \tag{5.1}$$

where $x = t^2$. The closed formula for $Q_{R,d}$ is given by the following theorem.

Theorem 5.2. *Let $Q_{R,d}$ and Q_R be formal Laurent series in $\mathbb{Q}((x))$ related by the formula* (5.1). *For any d and R we have*

$$Q_{R,d}(x) = \sum_{r\geq 1}\sum_{I}\frac{x^{M_R'(I;d)+M_R(I;(d+\alpha(R))/n(R))}}{(x^{n(R_1^I)+n(R_2^I)}-1)\cdots(x^{n(R_{r-1}^I)+n(R_r^I)}-1)}\prod_{k=1}^{r} Q_{R_k^I}(x), \tag{5.3}$$

where $M'_R(I;d)$ and $M_R(I;\lambda)$ for a partition I of R and $\lambda \in \mathbb{R}$ are defined by

$$M'_R(I;d) = -(n(R) - n(R^I_r))d - \sigma_R(I) + (2n(R) - n(R^I_1) - n(R^I_r)) \qquad \text{(5.4}$$

and

$$M_R(I;\lambda) = \sum_{k=1}^{r-1} (n(R^I_k) + n(R^I_{k+1}))[(n(R^I_1) + \cdots + n(R^I_k))\lambda - \alpha(R^I_{\leq k})]. \quad \text{(5.5}$$

Here $[x]$ for a real number x denotes the largest integer less than or equal to x.

Proof. As in Zagier [Z] we introduce a real parameter with respect to which we perform peculiar induction to prove the following theorem, which in turn implies theorem (5.2) b the substitution $\lambda = (d + \alpha(R))/n(R)$.

Theorem 5.6. *Let the hypothesis be as in the previous theorem. The two quantities*

$$Q^\lambda_{R,d}(x) = \sum_{r\geq 1} \sum_{I} \sum_{\circ_\lambda} x^{N_R(I;d_1,\ldots,d_r)} \prod_{k=1}^{r} Q_{R^I_k,d_k}(x), \qquad \text{(5.7}$$

$$S^\lambda_{R,d}(x) = \sum_{r\geq 1} \sum_{I} \frac{x^{M'_R(I;d)+M_R(I;\lambda)}}{(x^{n(R^I_1)+n(R^I_2)} - 1)\cdots(x^{n(R^I_{r-1})+n(R^I_r)} - 1)} \prod_{k=1}^{r} Q_{R^I_k}(x) \qquad \text{(5.8}$$

agree for every real number $\lambda \geq (d + \alpha(R))/n(R)$.

Here $\sum_{\circ_\lambda}$ denotes the summation over $(d_1, \ldots, d_r) \in \mathbb{Z}^r$ such that $\sum_i d_i = d$ and th following holds

$$\lambda \geq \frac{d_1 + \alpha(R^I_1)}{n(R^I_1)} > \frac{d_2 + \alpha(R^I_2)}{n(R^I_2)} > \cdots > \frac{d_r + \alpha(R^I_r)}{n(R^I_r)}. \qquad \text{(5.9}$$

Proof. We first note that $Q^\lambda_{R,d}$ and $S^\lambda_{R,d}$ are step functions of λ and they only jump at discrete subset of $\mathbb{R}$. We assume by induction that $Q^\lambda_{R',d} = S^\lambda_{R',d}$ for all data R' of ran $n(R') < n$, for all $d \in \mathbb{Z}$ and $\lambda \in \mathbb{R}$. Now for a given data R of rank $n(R) = n$ and $d \in$ we make the following claims:

Claim 1. Given any N, there exists $\lambda_0(N)$ such that for $\lambda \geq \lambda_0(N)$, $Q^\lambda_{R,d}$ and $S^\lambda_{R,d}$ agre modulo x^N.

Claim 2. For any $\lambda \geq (d + \alpha(R))/n(R)$, if we define $Q^{\lambda^-}_{R,d}$ (resp. $S^{\lambda^-}_{R,d}$) to be $Q^{\lambda-\epsilon}_{R,d}$ (res $S^{\lambda-\epsilon}_{R,d}$) for $\epsilon > 0$ small enough such that the function $Q^\lambda_{R,d}$ (resp. $S^\lambda_{R,d}$) has no jumps in t interval $[\lambda - \epsilon, \lambda)$, then the two functions $\Delta Q^\lambda_{R,d} = Q^\lambda_{R,d} - Q^{\lambda^-}_{R,d}$ and $\Delta S^\lambda_{R,d} = S^\lambda_{R,d} - S^{\lambda}_{R}$ are equal.

Proof of Claim 1. For a particular N, by equation (5.1), the coefficient of x^N in involves only finitely many choices of the integer r, partitions I, the integers $(d_1, \ldots, d$ Hence if we choose $\lambda_0(N) > (d_i + \alpha(R^I_i))/n_i$ for all such combinations of $(r, I, d_1, \ldots, d$ then the coefficient of x^N in Q_R and $Q^\lambda_{R,d}$ are equal for $\lambda \geq \lambda_0(N)$. On the other hand, $r > 1$, then the term $M_R(I;\lambda)$, occuring in the exponent of the numerator in eq. (5. tends to ∞ as λ tends to ∞. So, for a fixed N if we choose λ large enough, we do not any contribution for the coefficient of x^N in $S^\lambda_{R,d}$. But for $r = 1$ the part of the summati in $S^\lambda_{R,d}$ is just Q_R. Hence the Claim 1 follows.

Proof of Claim 2. Given a data R and a sub-data L of R we define a numerical function

$$\delta_R(L) = \sum_P \sum_{i>t} (R-L)_i^P L_t^P. \tag{5.10}$$

We first write down the recursions satisfied by the various numerical functions that we have encountered in the statement of the theorem (5.6).

Lemma 5.11. *Let I be a partition of R of length r. Let $0<k<r$.*

(a) $\sigma_R(I) = \sigma_{R^I_{\leq k}}(I_{\leq k}) + \sigma_{R^I_{\geq k+1}}(I_{\geq k+1}) + \delta_R(R^I_{\leq k})$

(b) $N_R(I; d_1, \ldots, d_r) - N_{R^I_{\geq 2}}(I_{\geq 2}; d_2, \ldots, d_r) = n(R^I_1)(n\lambda - d) - n\alpha(R^I_1) - \delta_R(R^I_1)$

(c)

$$\begin{aligned} &M'_{R^I_{\leq k}}(I_{\leq k}; d(\lambda, R^I_{\leq k})) + M'_{R^I_{\geq k+1}}(I_{\geq k+1}; d - d(\lambda, R^I_{\leq k})) \\ &\quad = M'_R(I; d) - (2n(R^I_{\leq k}) - n(R) - n(R^I_k) + n(R^I_r))d(\lambda, R^I_{\leq k}) \\ &\qquad - n(R^I_k) - n(R^I_{k+1}) + \delta_R(R^I_{\leq k}) + n(R^I_{\leq k})d. \end{aligned} \tag{5.12}$$

where $d(\lambda, L) = n(L)\lambda - \alpha(L)$ for any data L.

Proof. All these statements follow from straight forward calculations, so we will not give the details. □

We now compute $\Delta Q^\lambda_{R,d}$. It is zero unless there is a partition I of R and a r-tuple $(d_1, \ldots, d_r)$ with $\sum d_i = d$ such that $\lambda = (d_1 + \alpha(R^I_1))/n(R^I_1)$. For such a λ, we observe that

$$\Delta Q^\lambda_{R,d} = \sum_{r\geq 1} \sum_I \sum_{\substack{\circ\lambda \\ \lambda=(d_1+\alpha(R^I_1))/n(R^I_1)}} x^{N_R(I; d_1, \ldots, d_r)} \prod_{k=1}^r Q_{R^I_k, d_k}(x). \tag{5.13}$$

We can use the lemma (5.11) in the above formula and separate the expressions which have $k=1$ and $k \geq 2$. Hence the right hand side in the equation (5.13) becomes

$$\sum_{\substack{L \text{ sub-data of } R \\ d(\lambda, L) \in \mathbb{Z}}} x^{n(L)(n(R)\lambda - d) - n(R)\alpha(L) - \delta_R(L)} Q^\lambda_{L, d(\lambda, L)} Q^{\lambda^-}_{R-L, d-d(\lambda, L)}. \tag{5.14}$$

Now we compute $\Delta S^\lambda_{R,d}$ at a λ when there is a jump. This happens when $(n(R^I_1) + \cdots + n(R^I_k))\lambda - \alpha(R^I_k)$ is an integer for some partition I of R with $\ell(I) = r$ and for some positive integer $k < r$.

Fix a partition I of length r. Let

$$\pi_I = \{k < r \mid d(\lambda, R^I_{\leq k}) \in \mathbb{Z}\}. \tag{5.15}$$

One can see that $\Delta M(I; \lambda) = \sum_{k\in\pi_I} (n(R^I_k) + n(R^I_{k+1}))$ so

$$\begin{aligned} x^{M(I;\lambda)} - x^{M(I;\lambda^-)} &= x^{M(I;\lambda^-)} (x^{\sum_{k\in\pi_I}(n(R^I_k)+n(R^I_{k+1}))} - 1) \\ &= \sum_{k\in\pi_I} x^{M(I;\lambda^-) + \sum_{\{k'\in\pi_I \mid k'<k\}}(n(R^I_{k'})+n(R^I_{k'+1}))} (x^{(n(R^I_k)+n(R^I_{k+1}))} - 1) \\ &= \sum_{k\in\pi_I} x^{M(R^I_{\leq k}; \lambda) + M(R^I_{\geq k+1}; \lambda^-) + (2n(R) - 2n(R^I_{\leq k}) + n(R^I_k) - n(R^I_r))d(\lambda, R^I_{\leq k})} \\ &\qquad .(x^{(n(R^I_k)+n(R^I_{k+1}))} - 1). \end{aligned} \tag{5.16}$$

Let $g_R(I;d)$ denote the following rational function of x

$$\frac{x^{M'_R(I;d)+M_R(I;\lambda)}}{(x^{n(R^I_1)+n(R^I_2)}-1)\cdots(x^{n(R^I_{r-1})+n(R^I_r)}-1)}. \tag{5.17}$$

Using the lemma (5.11) and equation (5.16) we can verify that $\Delta(g_R(I;d))$ is equal to the following

$$\sum_{k\in\pi_I} x^{n(R^I_k)(n(R)\lambda-d)-n(R)\alpha(R^I_{\leq k})-\delta_R(R^I_{\leq k})}$$
$$\cdot g_{R^I_{\leq k}}(I_{\leq k}, d(\lambda, R^I_{\leq k}))g_{R^I_{\geq k+1}}(I_{\geq k+1}, d-d(\lambda, R^I_{\leq k})). \tag{5.18}$$

Now $\Delta S^\lambda_{R,d}$ is computed to be

$$\sum_{r\geq 1}\sum_{I}\Delta(g_R(I;d))\prod_{k=1}^{r}Q_{R^I_k}(x). \tag{5.19}$$

Using the equation (5.18), and grouping together all terms which give the sub-data L, we get the following expression for $\Delta S^\lambda_{R,d}$,

$$\sum_{L} x^{n(L)(n(R)\lambda-d)-n(R)\alpha(L)-\delta_R(L)}S^\lambda_{L,d_{\lambda,L}}S^{\lambda^-}_{R-L,d-d(\lambda,L)}, \tag{5.20}$$

where the summation is over sub-data L of R with $d(\lambda,L)\in\mathbb{Z}$. But $Q^\lambda_{L,d_{\lambda,L}}=S^\lambda_{L,d_{\lambda,L}}$ and $Q^{\lambda^-}_{R-L,d-d(\lambda,L)}=S^{\lambda^-}_{R-L,d-d(\lambda,L)}$ by induction (since $n(L)$ and $n(R)-n(L)$ are less than $n(R)$), hence we get $\Delta S^\lambda_{R,d}=\Delta Q^\lambda_{R,d}$. This proves claim 2.

To prove the theorem it is enough to check that the coefficient of x^N in $Q^\lambda_{R,d}$ and in $S^\lambda_{R,d}$ agree for any N. For a given N, the claim 1 implies that the coefficients of $Q^\lambda_{R,d}$ and $S^\lambda_{R,d}$ are equal when λ is sufficiently large. Since $Q^\lambda_{R,d}$ and $S^\lambda_{R,d}$ are step functions of λ jumping only at a discrete set of real numbers, and for such real numbers by claim 2 their jumps agree therefore the jumps in the coefficients also agree, which in turn proves that the coefficients are the same. This completes the proof of the theorem. □

Now if we define

$$\sigma'_R(I)=\sum_{P\in S}\sum_{k>l,i<t}I^P_{i,k}I^P_{t,l}, \tag{5.21}$$

then one observes that dimensions of the flag varieties $\mathcal{F}_R$ and $\mathcal{F}_{R^I_k}$ are related by

$$\dim\mathcal{F}_R-\sum_{k=1}^{r}\dim\mathcal{F}_{R^I_k}=\sigma_R(I)+\sigma'_R(I). \tag{5.22}$$

Using this expression we can formulate the closed formula for the Poincaré polynomial of the moduli space of parabolic semi-stable bundles as follows.

Theorem 5.23. *The Poincaré polynomial $P_{R,d}$ of the moduli space of parabolic stable bundles with a fixed determinant of degree d, and data R satisfying the condition 'par semi-stable = par stable' is given by*

$$\frac{1-t^2}{(1+t)^{2g}}\sum_{r\geq 1}\sum_{I}\frac{t^{2(\sigma'_R(I)-(n(R)-n(R^I_r))d+M_g(I;(d+\alpha(R))/n(R))}}{(t^{2n(R^I_1)+2n(R^I_2)}-1)\cdots(t^{2n(R^I_{r-1})+2n(R^I_r)}-1)}\prod_{k=1}^{r}P_{R^I_k}(t), \tag{5.24}$$

where $M_g(I;\lambda)$ *is*

$$\sum_{k=1}^{r-1}(n(R^I_k)+n(R^I_{k+1}))([(n(R^I_1)+\cdots+n(R^I_k))\lambda-\alpha(R^I_{\le k})]+1) + (g-1)\sum_{i<j}n(R^I_i)n(R^I_j) \tag{5.25}$$

and $P_R(t)$ *is defined to be*

$$\left(\frac{\prod_{i=1}^{n(R)}(1-t^{2i})^{|S|}}{\prod_{P\in S}\prod_{\{i|R^P_i\neq 0\}}\prod_{l=1}^{R^P_i}(1-t^{2l})}\right)\left(\frac{\prod_{i=1}^{n(R)}(1+t^{2i-1})^{2g}}{(1-t^{2n(R)})\prod_{i=1}^{n(R)-1}(1-t^{2i})^2}\right). \tag{5.26}$$

6. Sample calculations

Rank 2

Now we write down the Poincaré polynomial in more and more explicit forms for any data R such that $n(R)=2$.

Let T be a subset of S defined by $\{P\in S|R^P_1=1\}$, which is the set of vertices where the parabolic filtration is non-trivial. Then we get

$$P_R(t)=\frac{(1+t^2)^{|T|}(1+t)^{2g}(1+t^3)^{2g}}{(1-t^4)(1-t^2)}. \tag{6.1}$$

Given any partition I of R, we define a subset T_I of T by

$$T_I=\{P\in T|I^P_{1,1}=0\}. \tag{6.2}$$

From this definition we observe that $\sigma'_R(I)$ (as defined in (5.21)) is just $|T_I|$. Let $\chi_I : T\longrightarrow\{1,-1\}$ be defined by

$$\begin{aligned}\chi_I(P)&=1 \quad \text{if } P\in T_I\\ &=-1 \quad \text{otherwise.}\end{aligned} \tag{6.3}$$

Using the theorem (5.23) for rank 2 moduli we obtain the following.

PROPOSITION 6.4

For any degree d, the Poincaré polynomial for the moduli space of rank 2 *parabolic bundles with data R and satisfying the condition 'par semi-stable = par stable' is given by*

$$P_{R,d}(t)=\frac{(1+t^2)^{|T|}(1+t^3)^{2g}-(\sum_I t^{2(g+|T_I|+[\psi_I]+a_I)})(1+t)^{2g}}{(1-t^4)(1-t^2)}, \tag{6.5}$$

where

$$\psi_I=\sum_{P\in T}\chi_I(P)(\alpha^P_1-\alpha^P_2) \tag{6.6}$$

and a_I *is* 1 *or* 0 *depending on whether* $d+[\psi_I]$ *is even or odd.*

Now we put $g=0$ in the formula. Since $P_{R,d}$ is a power series in t, one sees that $|T_I|+[\psi_I]+a_I\ge 0$ for every partition I.

Using the proposition (6.4), the zeroth Betti number of the moduli space can b computed to be equal to $1 - |\{I||T_I| + [\psi_I] + a_I = 0\}|$, hence the quantity $|T_I| + [\psi_I] + a$ is 0 for at most one partition. Hence we obtain the following corollary.

COROLLARY 6.7

Assuming the condition 'par semi-stable = par stable' we have

(a) *The moduli space of parabolic semi-stable bundles of rank 2 is non-empty iff fo every partition I we have $|T_I| + [\psi_I] + a_I > 0$.*
(b) *The moduli is actually connected when it is non-empty.*

One can easily see that this condition is equivalent to the condition given by Biswas [B] Even in higher rank we can get a criterion for existence of stable bundles by settin $P_{R,d}(t) \neq 0$.

In what follows we assume that ψ_I is never an integer, which has the effect that th condition 'par semi-stable = par stable' holds for all the degrees.

Using the above formula for the Poincaré polynomial we compute it in an explicit form when the cardinality of S is small (1, 2, 3 and 4). For this, one observes that the abov expression for $P_{R,d}(t)$, the dependence on the weights is only via their differences. In vie of this we define $\delta^P = \alpha_1^P - \alpha_2^P$ for each $P \in S$.

When $S = \{P\}$, δ^P arbitrary, $R_i^P = 1$ for all i and any degree d, we compute th Poincaré polynomial to be

$$P_{R,d}(t) = \frac{(1+t^3)^{2g} - t^{2g}(1+t)^{2g}}{(1-t^2)^2}. \qquad (6.8$$

When $S = \{P_1, P_2\}$, δ^{P_1} and δ^{P_2} arbitrary, $R_i^P = 1$ for all i and P, and any degree d, we hav

$$P_{R,d}(t) = \frac{(1+t^2)((1+t^3)^{2g} - t^{2g}(1+t)^{2g})}{(1-t^2)^2}. \qquad (6.9$$

When $S = \{P_1, P_2, P_3\}$, $R_i^{P_j} = 1$ for all i and $j = 1, \ldots 3$, and any degree d. By reorderin P_1, P_2, P_3, we may assume that $\delta^{P_1} \leq \delta^{P_2} \leq \delta^{P_3}$. Now there are two possibilities:

(i) If $\delta^{P_1} + \delta^{P_2} + \delta^{P_3} < -2$ or $-\delta^{P_1} + \delta^{P_2} + \delta^{P_3} > 0$ then

$$P_{R,d}(t) = \frac{(1+t^2)^2((1+t^3)^{2g} - t^{2g}(1+t)^{2g})}{(1-t^2)^2}. \qquad (6.10$$

(ii) If $\delta^{P_1} + \delta^{P_2} + \delta^{P_3} > -2$ and $-\delta^{P_1} + \delta^{P_2} + \delta^{P_3} < 0$ (which is the remaining case), the

$$P_{R,d}(t) = \frac{(1+t^2)^2(1+t^3)^{2g} - 4t^{2g+2}(1+t)^{2g}}{(1-t^2)^2}. \qquad (6.1$$

When $S = \{P_1, P_2, P_3, P_4\}$, $R_i^{P_j} = 1$ for all i and j, and any degree d. Again we assun $\delta^{P_1} \leq \delta^{P_2} \leq \delta_3^P \leq \delta^{P_4}$. Again there are two possibilities:

(i) If $\delta^{P_1} + \delta^{P_2} + \delta^{P_3} - \delta^{P_4} < -2$ or $-\delta^{P_1} + \delta^{P_2} + \delta^{P_3} + \delta^{P_4} > 0$ then

$$P_{R,d}(t) = \frac{(1+t^2)^3((1+t^3)^{2g} - t^{2g}(1+t)^{2g})}{(1-t^2)^2}. \qquad (6.1$$

(ii) If $\delta^{P_1}+\delta^{P_2}+\delta^{P_3}-\delta^{P_4}>-2$ and $-\delta^{P_1}+\delta^{P_2}+\delta^{P_3}+\delta^{P_4}<0$ (which is the remaining case), then

$$P_{R,d}(t)=\frac{(1+t^2)^3(1+t^3)^{2g}-4t^{2g+2}(1+t^2)(1+t)^{2g}}{(1-t^2)^2}. \tag{6.13}$$

Remark 6.14. Note that in these cases we have considered, the Poincaré polynomial does not depend on the degree. In fact we can verify that in general for rank 2 the Poincaré polynomial is independent of the degree.

Ranks 3 and 4

Using the software *Mathematica*, we have computed the Poincaré polynomials and the Betti numbers for the rank 3 and rank 4 when the number of parabolic points is one or two. In this case we find that the Poincaré polynomial has dependence on the weights and degree. In the appendix we actually give the tables for the Betti numbers (in the rank 3 and rank 4 case) taking different set of weights into consideration. For the Poincaré polynomial we choose one sets of weights as an example in each of the following cases.

When rank = 3, $S=\{P_1\}$, $R_i^{P_1}=1$ for all i and assume that the condition 'par semi-stable = par stable' holds. Then for all choices of weights and degree d we have

$$\begin{aligned}P_{R,d}=\{&t^{6g-2}(1+t^2+t^4)(1+t)^{4g}-t^{4g-2}(1+t^2)^2(1+t)^{2g}(1+t^3)^{2g}\\&+(1+t^3)^{2g}(1+t^5)^{2g}\}/((t^2-1)^4(1+t^2)).\end{aligned} \tag{6.15}$$

When rank = 3, $S=\{P_1,P_2\}$, $R_i^{P_j}=1$ for all i and j. One observes that the condition 'par semi-stable = par stable' holds for all choices of degree. For $(\alpha_1^{P_1},\alpha_2^{P_1},\alpha_3^{P_1})=(0,1/12,3/12)$, $(\alpha_1^{P_2},\alpha_2^{P_2},\alpha_3^{P_2})=(1/12,5/12,6/12)$, one observes that the condition 'par semi-stable = par stable' holds for all choices of degree. When the degree $d=0$ or 2 mod 3 we find that

$$\begin{aligned}P_{R,d}=\{&-3t^{4g}(1+t)^{2g}(1+t^2)^2(1+t^3)^{2g}+t^{6g}(1+t)^{4g}(2+5t^2+2t^4)\\&+(1+t^3)^{2g}(1+t^5)^{2g}(1+t^2+t^4)\}/(1-t^2)^4\end{aligned} \tag{6.16}$$

and if $d=1$ mod 3 then the Poincaré polynomial is

$$\begin{aligned}P_{R,d}=(1+t^2+t^4)\{&t^{6g-2}(1+t^2+t^4)(1+t)^{4g}\\&-t^{4g-2}(1+t^2)^2(1+t)^{2g}(1+t^3)^{2g}\\&+(1+t^3)^{2g}(1+t^5)^{2g}\}/(1-t^2)^4.\end{aligned} \tag{6.17}$$

When rank is 4 and $|S|=1$ we find that the Poincare polynomial depends on the degree too. If we choose $R_i^P=1$ for all i, and choose $(\alpha_1^P,\alpha_2^P,\alpha_3^P,\alpha_4^P)=(0,1/8,1/4,1/2)$, then the condition 'par semi-stable = par stable' holds for all choices of degree. We find that

$$\begin{aligned}P_{R,0}=P_{R,1}=P_{R,2}=\{&(1+t^3)^{2g}(1+t^5)^{2g}(1+t^7)^{2g}\\&-2t^{-2+6g}(1+t)^{2g}(1+t^3)^{2g}(1+t^5)^{2g}(1+t^2+t^4)\\&-t^{-4+8g}(1+t)^{2g}(1+t^3)^{4g}(1+t^2+t^4)^2\\&+t^{-4+10g}(1+t^2)(1+t)^{4g}(1+t^3)^{2g}(3+5t^2+5t^4+3t^6)\\&-2t^{-4+12g}(1+t)^{6g}(1+t^2+t^4)^2\}/((1-t^2)^6(1+t^2)(1+t^2+t^4))\end{aligned} \tag{6.18}$$

and

$$\begin{aligned} P_{R,3} = \{&(1+t^3)^{2g}(1+t^5)^{2g}(1+t^7)^{2g} \\ &- t^{-4+6g}(1+t)^{2g}(1+t^3)^{2g}(1+t^5)^{2g}(1+t^2+t^4)(1+t^4) \\ &- t^{-4+8g}(1+t)^{2g}(1+t^3)^{4g}(1+t^2+t^4)^2 \\ &+ t^{-6+10g}(1+t^2)^4(1+t)^{4g}(1+t^3)^{2g}(1+t^4) \\ &-2t^{-6+12g}(1+t)^{6g}(1+t^4)(1+t^2+t^4)^2\}/((1-t^2)^6(1+t^2)(1+t^2+t^4)). \end{aligned} \tag{6.19}$$

If we choose the weights $(\alpha_1^P, \alpha_2^P, \alpha_3^P, \alpha_4^P) = (0, 1/5, 4/5, 9/10)$ then again the condition 'par semi-stable = par stable' holds for all choices of degree. If $P_{R,d}$ and $P'_{R,d}$ denote the Poincaré polynomial for the moduli space of parabolic stable bundles with data R (satisfying $n(R) = 4$), having degree d and with weights (0, 1/8, 1/4, 1/2) and (0, 1/5, 4/5, 9/10) respectively, then we find that $P'_{R,0} = P'_{R,2} = P_{R,0}$ and $P'_{R,1} = P'_{R,3} = P_{R,1}$.

7. Appendix : Betti number tables

The following tables give the Betti numbers up to the middle dimension of the moduli space of parabolic bundles over X for ranks 2, 3 and 4 and low genus. When $\beta_0 = 0$, we mean that the space is empty.

Rank 2

Any degree d, $R_i^P = 1$ for all i and $P \in S$.

Case A. $S = \{P_1\}$, δ^{P_1} arbitrary.
Case B. $S = \{P_1, P_2\}$, δ^{P_1} and δ^{P_2} arbitrary.
Case C. $S = \{P_1, P_2, P_3\}$, $\delta^{P_1} + \delta^{P_2} + \delta^{P_3} < -2$ or $-\delta^{P_1} + \delta^{P_2} + \delta^{P_3} > 0$
Case D. $S = \{P_1, P_2, P_3\}$, $\delta^{P_1} + \delta^{P_2} + \delta^{P_3} > -2$ and $-\delta^{P_1} + \delta^{P_2} + \delta^{P_3} < 0$.
Case E. $S = \{P_1, P_2, P_3, P_4\}$, $\delta^{P_1} + \delta^{P_2} + \delta^{P_3} - \delta^{P_4} < -2$ or $-\delta^{P_1} + \delta^{P_2} + \delta^{P_3} + \delta^{P_4} > 0$
Case F. $S = \{P_1, P_2, P_3, P_4\}$, $\delta^{P_1} + \delta^{P_2} + \delta^{P_3} - \delta^{P_4} > -2$ and $-\delta^{P_1} + \delta^{P_2} + \delta^{P_3} + \delta^{P_4} < 0$.

	Genus $g = 0$						Genus $g = 1$					
	A	B	C	D	E	F	A	B	C	D	E	F
β_0	0	0	0	1	0	1	1	1	1	1	1	1
β_1	–	–	–	–	–	0	0	0	0	0	0	0
β_2	–	–	–	–	–	–	–	2	3	4	4	5
β_3	–	–	–	–	–	–	–	–	0	2	0	2
β_4	–	–	–	–	–	–	–	–	–	–	6	8

	Genus $g = 2$						Genus $g = 3$					
	A	B	C	D	E	F	A	B	C	D	E	F
β_0	1	1	1	1	1	1	1	1	1	1	1	1
β_1	0	0	0	0	0	0	0	0	0	0	0	0
β_2	2	3	4	4	5	5	2	3	4	4	5	5
β_3	4	4	4	4	4	4	6	6	6	6	6	6
β_4	2	4	7	8	11	12	3	5	8	8	12	12

	Genus $g = 2$						Genus $g = 3$					
	A	B	C	D	E	F	A	B	C	D	E	F
β_5	–	8	12	16	16	20	12	18	24	24	30	30
β_6	–	–	8	14	15	22	18	21	26	27	34	35
β_7	–	–	–	–	24	32	12	24	42	48	66	72
β_8	–	–	–	–	–	–	–	36	57	72	83	99
β_9	–	–	–	–	–	–	–	–	48	68	90	116
β_{10}	–	–	–	–	–	–	–	–	–	–	114	144

Rank 3

Case A. $S = \{P\}$, $R_i^P = 1$ for all i. We take all choices of weights and degrees.

Case B. When $S = \{P_1, P_2\}$, $R_i^{P_1} = 1 = R_i^{P_2}$ for all i, $(\alpha_1^{P_1}, \alpha_2^{P_1}, \alpha_3^{P_1})= (0, 1/12, 3/12)$, $(\alpha_1^{P_2}, \alpha_2^{P_2}, \alpha_3^{P_2})= (1/12, 5/12, 6/12)$ $d = 0$ or $2 \bmod 3$.

Case C. $S=\{P_1, P_2\}$, $R_i^{P_1}=1=R_i^{P_2}$ for all i. $(\alpha_1^{P_1}, \alpha_2^{P_1}, \alpha_3^{P_1})=(0, 1/12, 3/12)$, $(\alpha_1^{P_2}, \alpha_2^{P_2}, \alpha_3^{P_2}) = (1/12, 5/12, 6/12)$, $d = 1 \bmod 3$.

	A, $g =$				B, $g =$				C, $g =$			
	0	1	2	3	0	1	2	3	0	1	2	3
β_0	0	1	1	1	0	1	1	1	0	1	1	1
β_1	–	0	0	0	–	0	0	0	–	0	0	0
β_2	–	2	3	3	–	5	5	5	–	4	5	5
β_3	–	0	4	6	–	2	4	6	–	0	4	6
β_4	–	–	7	7	–	12	15	15	–	8	15	15
β_5	–	–	16	24	–	6	24	36	–	0	24	36
β_6	–	–	18	28	–	16	40	49	–	10	39	49
β_7	–	–	36	60	–	–	80	120	–	–	76	120
β_8	–	–	45	103	–	–	108	176	–	–	98	176
β_9	–	–	56	140	–	–	188	314	–	–	164	314
β_{10}	–	–	70	261	–	–	251	531	–	–	203	530
β_{11}	–	–	64	354	–	–	344	784	–	–	264	778
β_{12}	–	–	–	537	–	–	436	1312	–	–	318	1293
β_{13}	–	–	–	780	–	–	480	1878	–	–	332	1828
β_{14}	–	–	–	998	–	–	528	2816	–	–	370	2697
β_{15}	–	–	–	1380	–	–	–	4036	–	–	–	3788
β_{16}	–	–	–	1652	–	–	–	5454	–	–	–	4983
β_{17}	–	–	–	1936	–	–	–	7442	–	–	–	6610
β_{18}	–	–	–	2170	–	–	–	9346	–	–	–	8007
β_{19}	–	–	–	2160	–	–	–	11526	–	–	–	9572
β_{20}	–	–	–	–	–	–	–	13394	–	–	–	10812
β_{21}	–	–	–	–	–	–	–	14562	–	–	–	11508
β_{22}	–	–	–	–	–	–	–	15210	–	–	–	11984

Rank 4

$|S| = 1$.

Case A. $R_i^P = 1$ for all i, $d = 0$ or 1 or 2 mod 4; $(\alpha_1^P, \alpha_2^P, \alpha_3^P, \alpha_4^P) = (0, 1/8, 1/4, 1/2)$ or $R_i^P = 1$ for all i, $d = 0$ or 2 mod 4, $(\alpha_1^P, \alpha_2^P, \alpha_3^P, \alpha_4^P) = (0, 1/5, 4/5, 9/10)$.

Case B. $R_i^P = 1$ for all i, $d = 3$ mod 4, $(\alpha_1^P, \alpha_2^P, \alpha_3^P, \alpha_4^P) = (0, 1/8, 1/4, 1/2)$ or $R_i^P = 1$ for all i, $d = 1$ or 3 mod 4, $(\alpha_1^P, \alpha_2^P, \alpha_3^P, \alpha_4^P) = (0, 1/5, 4/5, 9/10)$.

	A, g			B, g		
	0	1	2	0	1	2
β_0	0	1	1	0	1	1
β_1	–	0	0	–	0	0
β_2	–	4	4	–	3	4
β_3	–	2	4	–	0	4
β_4	–	8	11	–	5	11
β_5	–	4	20	–	0	20
β_6	–	10	31	–	6	31
β_7	–	–	64	–	–	64
β_8	–	–	90	–	–	89
β_9	–	–	164	–	–	160
β_{10}	–	–	241	–	–	232
β_{11}	–	–	376	–	–	356
β_{12}	–	–	563	–	–	521
β_{13}	–	–	792	–	–	712
β_{14}	–	–	1144	–	–	1001
β_{15}	–	–	1508	–	–	1272
β_{16}	–	–	2003	–	–	1635
β_{17}	–	–	2492	–	–	1952
β_{18}	–	–	2989	–	–	2263
β_{19}	–	–	3424	–	–	2528
β_{20}	–	–	3675	–	–	2660
β_{21}	–	–	3816	–	–	2760
β_{22}	–	–	–	–	–	–

Acknowledgements

The author wishes to thank Nitin Nitsure for his constant support and encouragement and for many useful comments and suggestions. The author also wishes to thank K P Yogendran for writing the computer program which was used to calculate the Betti numbers for ranks 3 and 4.

References

[A-B] Atiyah M F and Bott R, The Yang–Mills equations over Riemann surfaces. *Philos. Trans. R. Soc. London Series A* **308** (1982) 523–615

[B] Biswas I, A criterion for the existence of parabolic stable bundle of rank two over the projective line, to appear in the *Int. J. Math.* **9** (1998) 523–533

[D-R] Desale U V and Ramanan S, Poincaré polynomials of the variety of stable bundles, *Math. Ann.* **216** (1975) 233–244

[F-S] Furuta M and Steer B, Seifert-fibered homology 3-spheres and Yang–Mills equations on Riemann surfaces with marked points, *Adv. Math.* **96** (1992) 38–102

[G-L] Ghione F and Letizia M, Effective divisors of higher rank on a curve and the Siegel formula, *Composito Math.* **83** (1992) 147–159

[H-N] Harder G and Narasimhan M S, On the cohomology groups of moduli spaces of vector bundles over curves, *Math. Ann.* **212** (1975) 215–248

[M-S] Mehta V B and Seshadri C S, Moduli of vector bundles on curves with parabolic structures, *Math. Ann.* **248** (1980) 205–239

[N1] Nitsure N, Cohomology of the moduli of parabolic vector bundles, *Proc. Indian Acad. Sci.* **95** (1986) 61–77

[N2] Nitsure N, Quasi-parabolic Siegel formula. *Proc. Indian Acad. Sci.* **106** (1996) 133–137; Erratum: **107** (1997) 221–222 (alg-geom/9503001 on the Duke e-print server)

[S] Seshadri C S, Fibrés vectoriels sur les courbes algébriques, *Asterisque* **96** (1982)

[Z] Zagier Don, Elementary aspects of Verlinde formula and of the Harder–Narasimhan–Atiyah–Bott formula, *Israel Math. Conf. Proc.* **9** (1996) 445–462

Proc. Indian Acad. Sci. (Math. Sci.), Vol. 110, No. 3, August 2000, pp. 263–292.
© Printed in India

The algebra of G-relations

VIJAY KODIYALAM, R SRINIVASAN and V S SUNDER

Institute of Mathematical Sciences, C.I.T. Campus, Taramani, Chennai 600 113, India

MS received 30 November 1999

Abstract. In this paper, we study a tower $\{A_n^G(d) : n \geq 1\}$ of finite-dimensional algebras; here, G represents an arbitrary finite group, d denotes a complex parameter, and the algebra $A_n^G(d)$ has a basis indexed by 'G-stable equivalence relations' on a set where G acts freely and has $2n$ orbits.

We show that the algebra $A_n^G(d)$ is semi-simple for all but a finite set of values of d, and determine the representation theory (or, equivalently, the decomposition into simple summands) of this algebra in the 'generic case'. Finally we determine the Bratteli diagram of the tower $\{A_n^G(d) : n \geq 1\}$ (in the generic case).

Keywords. Equivalence relations; Bratteli diagrams; G-relations.

1. Introduction

Let R_n denote the set of equivalence relations on the set $[n] = \{1, 2, \ldots, n\}$, and let ρ_n denote the cardinality of R_n. By convention, $n = 0, 1, 2, \ldots$ and $\rho_0 = 1$. Easy counting arguments show that the first few values of the sequence $\{\rho_n : n \geq 0\}$ are given by 1, 1, 2, 5, 15, 52, 203, ..., and that the sequence satisfies the recursion relation

$$\rho_{n+1} = \sum_{k=0}^{n} \binom{n}{k} \rho_k, \quad \forall n \geq 0. \tag{1.1}$$

Given $P, Q \in R_n$, we shall say that $P \leq Q$ if any two P-related indices are necessarily Q-related – or equivalently, if every Q-equivalence class is a union of P-equivalence classes. Clearly, if $P_{\min}$ is the trivial relation all of whose equivalence classes are singletons, and if $P_{\max}$ is the equivalence relation with just one equivalence class, then $P_{\min} \leq P \leq P_{\max}$, $\forall$ P. It is not hard to see that R_n is a lattice with respect to this order structure. (For instance, if $n = 4$, $P = \{\{1,2\},\{3,4\}\}$ and $Q = \{\{1\},\{2,3,4\}\}$, then $P \vee Q = P_{\max}$ and $P \wedge Q = \{\{1\},\{2\},\{3,4\}\}$. We shall, as above, sometimes equate an equivalence relation with the set of its equivalence classes.)

Further, if $P \in R_n$, we shall write $||P||$ for the number of equivalence classes in P. Before proceeding further, we record a simple fact as a lemma, for convenience of future reference.

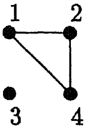
Lemma 1. *If* $P, Q \in R_n$, *then*,

(a) $||P \vee Q|| \leq ||P||$;
(b) *if* $||P \vee Q|| = ||P||$ *and* $P \neq Q$, *then* $||P|| < ||Q||$.

Proof. (a) Follows from the fact that every $P \vee Q$-equivalence class is a union of P-equivalence classes. (b) The hypothesis is seen to imply that no two indices which are

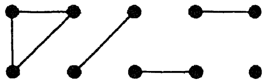
P-inequivalent can be $P \vee Q$-equivalent; this implies that $P \vee Q = P$. On the other han every Q-equivalence class is contained in a P-equivalence class, and the assumption th $P \neq Q$ says that at least one P-equivalence class must be the union of two or more equivalence classes, and the proof is complete.

We think of an element of R_{2n} as a diagram, thus: we think of the $2n$ elements of [2 listed in two rows of n elements each, with the jth point from the left on the top (resp bottom) row indexed by j (resp., $n+j$); and connect every pair of indices which a equivalent under the relation. For instance, the relation in R_4, whose equivalence class are $\{1,2,4\}$ and $\{3\}$ is represented by the picture

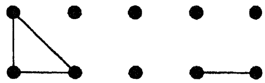


We will be interested in the vector space with R_{2n} as basis, which will be equipped wi the structure of a $\mathbb{C}$-algebra, with the definition of the product of basis vectors involving complex parameter d. Rather than giving a precise and rigorous definition, we sh describe the prescription for an example.

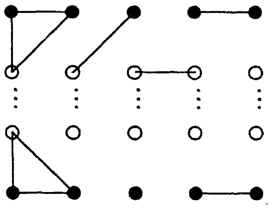
For instance, suppose $n = 5, P = \{\{1,2,6\},\{3,7\},\{4,5\},\{8,9\},\{10\}\}$ and Q $\{\{1,6,7\},\{2\},\{3\},\{4\},\{5\},\{8\},\{9,10\}\}$; according to our diagrammatic notatio we have:

$P =$ 

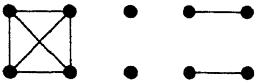
and

$Q =$ 

In order to define the product PQ, first concatenate the pictures (with P on top and Q the bottom), and identify the intermediate levels of points as indicated:



then introduce a power of d equal to the number of 'components' in the grand pict which are entirely contained in the two middle levels, and then forget the two mid levels altogether, to finally obtain:

$PQ = d^2 \cdot$ 

It is relatively painless to verify that this definition yields a finite-dimensional associative $\mathbb{C}$-algebra (of dimension ρ_{2n}), which we denote by $A_n(d)$. This algebra has a multiplicative identity, i.e., the equivalence relation which has n equivalence classes, namely $\{k, n+k\}, 1 \leq k \leq n$.

As a trivial example, $A_1(d)$ has a basis consisting of 2 elements–say $1 = \{\{1,2\}\}$ and $P = \{\{1\},\{2\}\}$ – where 1 is the multiplicative identity, and $P^2 = dP$.

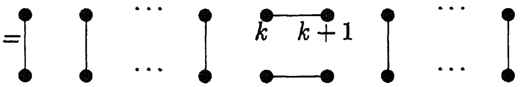
The process of 'adding a single vertical line at the right extreme' yields an injective map from R_{2n} into R_{2n+2}, which is easily seen to linearly extend to a multiplicative (identity-preserving) homomorphism of $A_n(d)$ into $A_{n+1}(d)$; further, since this map sends a basis injectively into a basis, it is necessarily a monomorphism. We thus have a tower

$$A_1(d) \subset A_2(d) \subset \cdots A_n(d) \subset \cdots$$

of finite-dimensional $\mathbb{C}$-algebras.

Remark 2. The tower $\{A_n(d)\colon n \geq 0\}$ has several interesting subtowers.

(a) The Temperley–Lieb algebra: Consider the subalgebra $T_n(d)$ of $A_n(d)$ consisting of the linear span of those equivalence relations P which satisfy two conditions: (i) each P-equivalence class contains precisely two elements; and most importantly, (ii) P admits a diagram–as in the above discussions–which is *planar*, i.e., the diagram has no crossings and is a planar diagram contained in the rectangle bounded by the $2n$ points. It is clear that the inclusion of $A_n(d)$ into $A_{n+1}(d)$ maps $T_n(d)$ into $T_{n+1}(d)$; and it is a fact that $T_n(d)$ is generated as a unital algebra by the elements $P_1, P_2, \ldots, P_{n-1}$, where, for $1 \leq k < n$,

$P_k =$ 

It is to be noted that $P_k^2 = dP_k$ and that $P_k P_{k\pm 1} P_k = P_k$; so, if we define $e_k = d^{-1}P_k$, then the e_k's are idempotents which satisfy $e_k e_{k\pm 1} e_k = d^{-2} e_k$.

(b) The group algebra $\mathbb{C}\Sigma_n$ of the symmetric group sits naturally as a subalgebra of $A_n(d)$ as follows: given $\sigma \in \Sigma_n$, let P_σ denote the equivalence relation, whose equivalence classes are $\{\{\sigma(k), k+n\} : 1 \leq k \leq n\}$. It is fairly easy to verify that this map is multiplicative, meaning that $P_\sigma \cdot P_\tau = P_{\sigma\tau}$. The following little observation, which we call (c) below for the sake of future reference, is proved easily.

(c) The following conditions on an element $P \in R_{2n}$ are equivalent: (i) P is an invertible element of $A_n(d)$; (ii) there exists a (necessarily unique) permutation $\sigma \in \Sigma_n$ such that $P = P_\sigma$, as in (b) above; (iii) P has precisely n 'through classes' in the sense of Definition 3 below.

DEFINITION 3

If $P \in R_{2n}$, a *through class of P* is an equivalence class A of P such that $A \cap \{1,2,\ldots,n\} \neq \emptyset$ and $A \cap \{n+1, n+2, \ldots, 2n\} \neq \emptyset$. We write $t(P)$ for the number of through classes of P.

For instance, in the preceding example illustrating the definition of the product, we have $t(P) = 2$ and $t(Q) = t(d^{-2} \cdot PQ) = 1$. This is an instance of a general fact which has consequences for much of the following discussion.

Lemma 4. *Suppose* $P, Q \in R_{2n}$, *and* $P \cdot Q = d^r \cdot S$, *for some* r *and some* $S \in R_{2n}$. *Ther*

$$t(S) \leq \min\{t(P), t(Q)\}.$$

Proof. This follows easily from the definitions. □

COROLLARY 5

For $0 \leq k \leq n$, *define* I_k^n *to be the set of those equivalence relations with exactly through classes*; *and let* I_k *be the linear subspace spanned by* $\cup_{r \leq k} I_r^n$; *then*

$$\{0\} = I_{-1} \subset I_0 \subseteq \cdots \subseteq I_k \subseteq I_{k+1} \subseteq \cdots \subseteq I_n = A_n(d) \qquad (1.$$

is a filtration of $A_n(d)$ *by two-sided ideals.*

Proof. Obvious. □

Before concluding this introduction, we shall briefly dwell on the manner in which v would like to think of elements of R_{2n}, viz., as consisting of three pieces of informatio (a) the 'top', (b) the data on how the top is connected to the bottom, and (c) the 'bottom

Thus, suppose $P \in R_{2n}$ and $t(P) = k$. Focus attention first on the top set of n points the diagram representing P; we can naturally associate an element $P^+ \in R_n$, together wi an unordered collection $\{C_j : 1 \leq j \leq k\}$ of 'distinguished' P^+-equivalence classes–the corresponding precisely to the intersections of through-classes of P and the top line. Th is the motivation for the following definition.

DEFINITION 6

In the sequel, the symbol S_k^n will denote the set of symbols of the form $\underline{R} = (R; \{C_j(\underline{R}$ $1 \leq j \leq k\})$, where $R \in R_n$ and $\{C_j(\underline{R}) : 1 \leq j \leq k\}$ is an unordered collection of distinct 'distinguished' R-equivalence classes. (Note, in particular, that if $\underline{R} \in S_k^n$, th $||R|| \geq k$.)

We shall want to encode an element P of R_{2n}, for which $t(P) = k$, as a triple $(\underline{P^+}, \rho, \underline{P}$ where $\underline{P^\pm} \in S_k^n$, and ρ is a permutation, in such a way that (i) the P-equivalence class which are entirely contained in $\{1, 2, \ldots, n\}$ are the same as the P^+-equivalence class other than $C_j(\underline{P^+})$, $1 \leq j \leq k$; (ii) the P-equivalence classes which are entirely contain in $\{n+1, n+2, \ldots, 2n\}$ are the same as the sets $(n + C) = \{n + m : m \in C\}$ where C a P^--equivalence class other than the $C_j(\underline{P^-})$, $1 \leq j \leq k$; and (iii) the k through-classes P are given by $C_j = C_{\rho(j)}(P^+) \cup (n + C_j(\underline{P^-}))$, $\forall\ 1 \leq j \leq k$. (In order to make prec sense of the permutation ρ, we should first choose some 'canonical ordering' of t collection of distinguished classes for each element of S_k^n; we will say no more on t here, since we will elaborate on it later.)

We record a lemma (which is a consequence of the definitions) for later reference; omit the simple proof.

Lemma 7. *Suppose* $P_1 \cdot P_2 = d^l Q$, *with* $P_j, Q \in R_{2n}$ *and* $l \in \mathbb{Z}_+$.

(a) *Assume that* $t(P_1) = t(Q)$. *Then*, $\underline{P_1^+} = \underline{Q^+}$.
(b) *Dually, if* $t(P_2) = t(Q)$, *then* $\underline{P_2^-} = \underline{Q^-}$.

After this paper was written up, the authors discovered that the algebras $A_n(d)$ h been extensively studied by Paul Martin–see [M, M1]; he calls them *partition algeb* and even discusses their representation theory in the 'non-generic case'. We, on the other hand, discuss only the case of 'generic d' when the algebras are semisimple, but we consider the 'equivariant case'. Specifically, for any finite group G, we consider an algebra $A_n^G(d)$–which has a basis of 'G-stable equivalence relations'–and show (Theorem 21) that these algebras are 'generically semisimple' and obtain (Theorem 33) the Bratteli diagram for the tower $\{A_n^G(d) : n \geq 1\}$.

2. *G*-relations

We shall now consider an 'equivariant version' of the above analysis. (We should perhaps mention that one of the reasons for our study of these algebras is the hope that we might be able to tie them up with the theory of 'planar algebras' developed by Jones (see [J1]); this will be discussed elsewhere.)

We begin with some notation. For a set X, we shall write $R(X)$ for the set of all equivalence relations on X–so that $R([n])$ is what was denoted by R_n in the last section. Suppose now that a group G acts on the set X; clearly then, we have a natural action of G on $R(X)$–given by $g \cdot R = \{(g \cdot x, g \cdot y) : (x,y) \in R\}$ whenever $R \in R(X)$ and $g \in G$. Call a relation *G-stable* if it is fixed by every element of G, and let $R^G(X)$ denote the set of all G-stable equivalence relations.

We shall only consider the case when G acts freely on X; when G and X are finite, the case we shall be concerned with, this amounts to assuming that $X = G \times \{1, 2, \ldots, n\}$ and that the action is defined by $g \cdot (x, i) = (g \cdot x, i)$. We shall denote this set by X_n in the sequel.

First consider the case $n = 1$. In the following lemma and elsewhere in this paper, the symbol $\coprod$ always denotes 'disjoint union'.

Lemma 8. (a) *If H is a subgroup of G, then the partition $G = \coprod g_i H$ of G into distinct left H-cosets yields a G-relation on X_1.* (b) *Every G-relation on X_1 arises as in* (a) *above. (Thus, there is a natural bijection between $R^G(X_1)$ and the set of subgroups of G.)*

Proof. (a) is clear; as for (b), let H denote the equivalence class of 1 (the identity of G) with respect to a G-stable relation. Suppose $h_1, h_2 \in H$; then, $1 \sim h_1 \Rightarrow h_1^{-1} \sim h_1^{-1} h_1 = 1$, and $h_1 h_2 \sim h_1 \sim 1$; so H is a subgroup. □

DEFINITION 9

Let $C = C(G)$ denote a collection of subgroups of G containing exactly one subgroup from each conjugacy class of subgroups; for each $H \in C$, let $N(H)$ be the normaliser in G of H, and suppose

$$G = \coprod_{\kappa} N(H) \sigma_{\kappa}^{H}, \tag{2.3}$$

where we always assume that $\{\sigma_\kappa^H : 1 \leq \kappa \leq [G : N(H)]\}$ contains the identity element G. (Thus, we have chosen fixed coset-representatives for $N(H) \backslash G$, choosing the identity as the representative of the coset $N(H)$.)

Remark 10. (a) Suppose $P \in R^G(X_n)$. Then the equation

$$R^P = \{(i,j) : 1 \leq i, j \leq n, \exists g, h \in G \text{ such that } ((g,i),(h,j)) \in P\}$$

defines an element of $R([n])$; it follows from the definition that $[(g,i)]_P$ is a through-cl
of P if and only if $[i]_{R^P}$ is a through-class of R^P (provided n is even, so that this mal
sense). (As above, we shall always use the notation $[i]_R$ to denote the R-equivalence cl
of the point i.) (b) For each $i \in [n]$, it follows from lemma 8 that there exists a unic
subgroup, say $K_i(P)$, such that $((g,i),(h,i)) \in P \Leftrightarrow g \in hK_i(P)$. (c) If $((g,i),(h,j)) \in P$
as in (a), then $gK_i(P)g^{-1} = hK_j(P)h^{-1}$, and in particular, the subgroups $K_i(P)$ and K_j
are conjugate whenever $(i,j) \in R^P$. (Reason: fix $k_j \in K_j(P)$; then $[(1,i)]_P = [(g^{-1}h, j)]$
$[(g^{-1}hk_j, j)]_P$; similarly $[(1,j)]_P = [(h^{-1}g, i)]_P$, and hence $[(1,i)]_P = [(g^{-1}hk_jh^{-1}g, i$
i.e., $g^{-1}hK_j(P)h^{-1}g \subseteq K_i(P)$; the reverse inclusion follows identically.) (d) Thus, e
$P \in R^G(X_n)$ determines a function

$$[n]/R^P \ni C \mapsto H_C^P \in C,$$

where H_C^P is the unique element of C which is conjugate to $K_i(P)\ \forall i \in C$. Further, for e
$C \in [n]/R^P$, we shall consistently use the notation $i(C) = \min\{i : i \in C\}$.

PROPOSITION 11

Let $P \in R^G(X_n)$ and R^P be as above. Then,

(a) *there exists a unique function $\phi^P : [n] \to \coprod_{H\in C} H\backslash G$, which satisfies the follow
conditions for all R^P-equivalence classes C:*

(i) $\phi^P(i) \in H_C^P\backslash G\ \forall i \in C$;

(ii) $\cup_{i\in C}(\phi^P(i) \times \{i\})$ *is a P-equivalence class*; (*here, we think of an elemen
$H\backslash G$ naturally as a subset of G*); *and*

(iii) $\phi^P(i(C)) = H_C^P\sigma_C^P$, *where* $\sigma_C^P \in \{\sigma_\kappa^{H_C^P} : 1 \leq \kappa \leq [G : N(H_C^P)]\}$.

(b) *Conversely, suppose we are given (i) an $R \in R([n])$, (ii) a map $[n]/R \ni C \mapsto H_C \in$
and (iii) a map $\phi : [n] \to \coprod_{H\in C} H\backslash G$, which satisfy*:

(i)′ $\phi(i) \in H_C\backslash G$, *whenever i belongs to the R-equivalence class C; and*

(iii)′ $\phi(i(C)) = H_C\sigma_C$, *where* $\sigma_C \in \{\sigma_\kappa^{H_C} : 1 \leq \kappa \leq [G : N(H_C)]\}\forall C$.

*Then, there exists a unique $P \in R^G(X_n)$ such that $R^P = R, H_C = H_C^P\ \forall C$, and $\phi^P =$
Further, this relation P is defined by*

$$((g,i),(h,j))) \in P \Leftrightarrow (i,j) \in R \text{ and } \phi(i)g^{-1} = \phi(j)h^{-1}. \qquad ($$

(c) *If ψ^P is another function defined on $[n]$ and satisfying (a) (i), (ii), then for each
equivalence class C, there exists a unique element $\omega_C^P \in N(H_C)/H_C$ such
$\psi^P(i) = \omega_C^P\phi^P(i)\ \forall i \in C$.*

Proof. (a) We first discuss uniqueness. Suppose we are given a function ϕ^P satisf
conditions (a) (i)–(iii). These conditions, and the definition of $K_i(P)$ shows th
$[i]_{R^P} = C$, then $\phi^P(i)$ is a left-coset of $K_i(P)$ as well as a right-coset of H_C^P. Sup
$\phi^P(i) = H_C^P g = g_1K_i(P)$; then clearly $g^{-1}H_C^P g = (\phi^P(i))^{-1}\phi^P(i) = K_i(P)$. In partic
this is true for $i = i(C)$, and since the σ_i^H are representatives of the distinct right-cose
$N(H)$, it follows that there exists a unique σ_C^P satisfying condition (iii). Now if we de
$D = [(\sigma_C^P, i(C))]_P$, we see from condition (ii) that for each $i \in C$, we must
$\phi^P(i) \times \{i\} = D \cap (G \times \{i\})$. This proves that the function ϕ^P is uniquely determine
the conditions (i)–(iii).

For existence, let us define σ_C^P and ϕ^P by the prescription forced by the discussion of last paragraph. We only need to verify that $\phi^P(i)$ is a right-coset of H_C. What is clear from the definition is that $\phi^P(i)$ is a left $K_i(P)$-coset; on the other hand, notice that $\phi^P(i(C))$ is invariant under the action of H_C^P, and that this is necessarily true also of D, and hence of $\phi^P(i)$, for each $i \in P$; thus, $\phi^P(i)$ is a left $K_i(P)$-coset, as also a union of right H_C^P-cosets; for reasons of cardinality, this forces $\phi^P(i)$ to be exactly one right H_C^P-coset, as desired. This proves existence.

(b) If the data of (b) (i)–(iii) satisfies (i)′, (iii)′, then equation (2.4) defines a G-stable equivalence relation. (Reason: the P-equivalence classes are just the 'sets of constancy' for the function $(g,i) \mapsto ([i]_R, H_{[i]_R}g^{-1})$.) The definition of P and of R^P implies that $R^P \subseteq R$; conversely, suppose $(i,j) \in R$; let $C = [i]_R = [j]_R$; since G acts transitively on $H_C\backslash G$, we can find $g \in G$ such that $\phi(i)g^{-1} = \phi(j)$; hence $((g,i),(1,j)) \in P$; this implies that $(i,j) \in R^P$. Thus, indeed $R = R^P$.

Let C be an R^P-equivalence class. By condition (iii)′, we have $\phi(i(C))\sigma_C^{-1} = H_C$, and hence, by definition of P,

$$\begin{aligned}((\sigma_C, i(C)), (g,j)) \in P &\Leftrightarrow j \in C \text{ and } H_C = \phi(j)g^{-1}\\ &\Leftrightarrow j \in C \text{ and } \phi(j) = H_C g\\ &\Leftrightarrow j \in C \text{ and } g \in \phi(j),\end{aligned}$$

(since $\phi(j)$ is given to be a right-coset of H_C for $j \in C$). Hence ϕ also satisfies:

(ii)′ $D = \cup_{i\in C}(\phi(i) \times \{i\})$ is a P-equivalence class.

Then, for any $i \in C$, it follows from the definition of $K_i(P)$ that $\phi(i)$ is a left-coset of $K_i(P)$ as well as a right-coset of H_C; this means that the subgroups $K_i(P)$ and H_C are conjugate whenever $i \in C$. Thus, we see that $H_C^P = H_C \ \forall C$.

So, the function ϕ satisfies the conditions (a) (i)–(iii), and we deduce from the uniqueness assertion of (a) that $\phi = \phi^P$.

(c) If we set $D = \cup_{i\in C}(\psi^P(i) \times \{i\})$, we see as in the proof of (b) above that if C is any R^P-equivalence class, then $\psi^P(i(C))$ is a left-coset of $K_{i(C)}(P)$ as well as a right-coset of H_C^P; if $\psi^P(i(C)) = H_C^P g$, this means that $K_{i(C)} = g^{-1}H_C^P g$. We already know that $K_{i(C)} = (\sigma_C^P)^{-1}H_C^P\sigma_C^P$. This means that $g(\sigma_C^P)^{-1} \in N(H_C^P)$ and hence there exists a unique element $\omega_C^P \in N(H_C^P)$ such that $g = \omega_C^P\sigma_C^P$. The definitions show that

$$\psi^P(i(C)) = H_C^P g = H_C^P\omega_C^P\sigma_C^P = \omega_C^P H_C^P\sigma_C^P = \omega_C^P\phi^P(i(C)).$$

It is now easy to verify that the function defined by $\phi(i) = (\omega^P_{[i]_{R^P}})^{-1}\psi^P(i)$ satisfies the three conditions (a) (i)–(iii), and an appeal to the uniqueness assertion of (a) completes the proof. □

Notice now that for any positive integer n, we may regard $R^G(X_n)$ as a subset of $R([n|G|])$; furthermore, if $P, Q \in R^G(X_{2n})$, and if $P \cdot Q = d^l S$, where the product is computed as in the algebra $A_{n|G|}(d)$, then it is easy to see that S corresponds to a G-stable equivalence relation on X_{2n}. Thus, the linear span of $R^G(X_{2n})$ is a subalgebra of $A_{2n|G|}(d)$.

DEFINITION 12

Let $A_n^G(d)$ denote the (finite-dimensional) algebra, with basis $R^G(X_{2n})$, obtained as above.

Remark 13. Let $P \in R^G(X_{2n})$; since G acts transitively on each $G \times \{j\}$, it is seen that if C is any through-class of R^P, then $G \times C$ is the disjoint union of $[G : H_C^P]$ many P-equivalence classes; and, as C varies over the through-classes of R^P, these exhaust all the through-classes of P; hence, if $t(P) = k$, then

$$k = \sum [G : H_C^P], \tag{2.5}$$

where the sum is over all through-classes C of R^P; in particular, if $t = t(R^P)$, then,

$$t \leq k \leq t|G|.$$

DEFINITION 14

(a) For $0 \leq k \leq n|G|$, define I_k^G to be the linear subspace of $A_n^G(d)$ spanned by $\{P \in R^G(X_{2n}) : t(P) \leq k\}$.

(b) If $P \in R^G(X_{2n})$, define $n^P : \mathcal{C} \to \mathbb{Z}_+ (= \{0, 1, \cdots\})$ by $n_P(H) = \#\{C : C$ is a through-class of R^P such that $H_C^P = H\}$ for all $H \in \mathcal{C}$.

(c) For $0 \leq k \leq n|G|$, let N_k denote the set of functions $\bar{n} : \mathcal{C} \to \mathbb{Z}_+$ which satisfy the conditions $\sum_H \bar{n}(H) \leq n$, and $k = \sum_H \bar{n}(H)[G : H]$. (Later, when we wish to vary n, we shall denote this object by the symbol $N_{n;k}$, since the definition also involves the inequality depending upon n.)
Let $N_{[n]} = \cup_{k=0}^{n|G|} N_k$.

(d) For arbitrary $\bar{n} \in N_{[n]}$, define $I(\bar{n}) = \{P \in R^G(X_{2n}) : n_P = \bar{n}\}$.

Thus, as in Corollary 5, it is true that $\{I_k^G : 0 \leq k \leq n|G|\}$ is a filtration of $A_n^G(d)$ by two-sided ideals.

Lemma 15. *For $0 \leq k \leq n|G|$ and arbitrary $\bar{n} \in N_k$, let $Q(\bar{n})$ denote the linear subspace spanned by $\pi(I(\bar{n}))$, where $\pi : I_k^G \to I_k^G / I_{k-1}^G$ is the quotient map; then, $Q(\bar{n})$ is an ideal in I_k^G / I_{k-1}^G, and further,*

$$I_k^G / I_{k-1}^G = \oplus_{\bar{n} \in N_k} Q(\bar{n}).$$

Proof. The lemma is a tautology when $k = 0$, so we may assume $k > 0$.

It should be clear that it is sufficient to prove that if $P_j \in I(\bar{n}_j), \bar{n}_j \in N_k, j = 1, 2,$ i $P_1 \cdot P_2 = d^l Q$ in $A_n^G(d)$, and if $t(Q) = k$, then $\bar{n}_1 = \bar{n}_2 = n_Q$.

In view of Lemma 7, it suffices to observe that n_P is uniquely determined by $\underline{P^+}$ as wel as by $\underline{P^-}$ – and this follows easily from the definitions. □

In order to arrive at a 'working description' of elements of these ideals, we shall firs obtain an alternative way of encoding the 'tops' of elements $P \in R^G(X_{2n})$. On the on hand, we can forget that P is G-stable and represent the 'top' and 'bottom' of P, and jus look at what we denoted by $\underline{P^\pm}$ at the end of § 1. Thus, for instance $\underline{P^+}$ is just the data P of the equivalence relation obtained by restricting P to the top (i.e., $G \times [n]$), togethe with the data of which P^+-equivalence classes are contained in through-classes of P.

We wish to bring in the knowledge of G-invariance of P to encode this data differentl For this, the starting point is the observation – see Remark 13 – that through-classes of are intimately tied with through-classes of R^P. We begin by trying to list the elements c the latter collection in a 'canonical order'.

If $n_P = \bar{n}$, and if $H \in \mathcal{C}$, then there exist $\bar{n}(H)$ many 'distinguished' R^{P^+}-equivalenc classes C^+ for which $H_{C^+}^{P^+} = H$; let $\{C_{H,s}(\mathbf{P}^+) : 1 \leq s \leq \bar{n}(H)\}$ be the unique listing c these classes which satisfies

$$s < s' \Rightarrow i(C_{H,s}(\mathbf{P}^+)) < i(C_{H,s'}(\mathbf{P}^+)). \tag{2.6}$$

DEFINITION 16

For $\overline{n} \in N_{[n]}$, define $S(\overline{n})$ to be the collection of all symbols $\mathbf{P}^+ = (P^+; \{C_{H,s}(\mathbf{P}^+) : 1 \leq s \leq \overline{n}(H), H \in \mathcal{C}\})$, where $P^+ \in R^G(X_n)$, and $\{C_{H,s}(\mathbf{P}^+) : 1 \leq s \leq \overline{n}(H), H \in \mathcal{C}\}$ is a collection of 'distinguished' R^{P^+}-equivalence classes such that (i) $H^{P^+}_{C_{H,s}(\mathbf{P}^+)} = H$ for all H, s, and (ii) the condition (2.6) is satisfied.

Thus, if $P \in R^G(X_{2n})$, and if $n_P = \overline{n}$, then the 'top' (resp., the 'bottom') of P determines an element $\mathbf{P}^+$ (resp., $\mathbf{P}^-$) of $S(\overline{n})$. Conversely, this $\mathbf{P}^+$ uniquely determines all the 'distinguished' classes of what we earlier called $\underline{P^+}$, since a P^+-equivalence class, say D^+, is contained in a through-class for P if and only if there exists a through-class, say C, of R^P such that $D^+ \subset (G \times C)$. Thus, what we have called $\mathbf{P}^+$ is nothing but another way of encoding what was earlier called $\underline{P^+}$ in case P is G-stable. Thus, in future, we shall freely use such expressions as 'let $\mathbf{P}^\pm$ denote the 'top' and 'bottom' of $P \in R^G(X_{2n})$'.

Lemma 17. *There exists a bijection*

$$I(\overline{n}) \ni P \overset{\zeta}{\mapsto} (\mathbf{P}^+, \rho(P), \mathbf{P}^-) \in S(\overline{n}) \times G(\overline{n}) \times S(\overline{n}),$$

where (*i*) $G(\overline{n}) = \prod_{H \in \mathcal{C}}((N(H)/H)^{\overline{n}(H)} \rtimes \Sigma_{\overline{n}(H)})$ *is the product* (*over the H's*) *of semi-direct-products* (*with respect to the natural permutation action of the second factor on the first*), *and* (*ii*) $\mathbf{P}^\pm$ *denote the 'top' and 'bottom' of P.*

Proof. Fix a $P \in I(\overline{n})$. For $Q \in \{P, P^+, P^-\}$, let ϕ^Q be the function associated to Q as in Proposition 11. By considering the through-classes of R^P, it is not hard to see that, for each fixed $H \in \mathcal{C}$, there is a unique permutation $\gamma_H \in \Sigma_{\overline{n}(H)}$ such that $\{(n + C_{H,s}(\mathbf{P}^-)) \cup C_{H,\gamma_H(s)}(\mathbf{P}^+) : 1 \leq s \leq \overline{n}(H)\}$ is precisely the collection of those R^P-through classes C for which $H^P_C = H$.

Notice next that the function defined on $[n]$ by $\psi^{P^-}(j) = \phi^P(n+j)$, satisfies the conditions of Proposition 11(c) (with P^- in place of the P there). Hence, by that proposition, for each R^{P^-}-equivalence class C^-, there exists a unique element $\omega^{P^-}_{C^-} \in N(H^{P^-}_{C^-})/H^{P^-}_{C^-}$ such that $\psi^{P^-}(j) = \omega^{P^-}_{C^-}\phi^{P^-}(j) \forall j \in C^-$. Set $\omega^H_s = \omega^{P^-}_{C_{H,\gamma_H^{-1}(s)}(\mathbf{P}^-)}$, for $1 \leq s \leq \overline{n}(H), H \in \mathcal{C}$.

Now define $\rho(P) = ((\rho(P)_H))_{H \in \mathcal{C}}$, where $\rho(P)_H \in (N(H)/H)^{\overline{n}(H)} \times \Sigma_{\overline{n}(H)}$ is defined by

$$\rho(P)_H = ((\omega^H_1, \ldots, \omega^H_{\overline{n}(H)}), \gamma_H).$$

Thus, we have defined the map ζ.

Conversely, suppose the triple $(\mathbf{P}^+, \rho(P), \mathbf{P}^-)$ is given, and suppose $\rho(P) = ((\rho(P)_H))_{H \in \mathcal{C}}$, where $\rho(P)_H = ((\omega^H_1, \ldots, \omega^H_{\overline{n}(H)}), \gamma_H)$. Then define:

(i) a relation $R \in R([2n])$ by demanding that its equivalence classes are: (a) the R^{P^+}-equivalence classes other than the $C_{H,s}(\mathbf{P}^+)$'s; (b) sets of the form $(n + C)$, where C is an R^{P^-}-equivalence class other than the $C_{H,s}(\mathbf{P}^-)$'s; and (c) $\{(n + C_{H,s}(\mathbf{P}^-)) \cup C_{H,\gamma_H(s)}(\mathbf{P}^+)) : 1 \leq s \leq \overline{n}(H), H \in \mathcal{C}\}$;

(ii) a map $[2n]/R \to \mathcal{C}$ by setting H_C to be equal to: (a) $H^{P^+}_C$, if C is an R^{P^+}-equivalence class other than the $C_{H,s}(\mathbf{P}^+)$'s; (b) $H^{P^-}_{C-n}$, if $(C-n)$ is an R^{P^-}-equivalence class other than the $C_{H,s}(\mathbf{P}^-)$'s; and (c) H if $C = (n + C_{H,s}(\mathbf{P}^-)) \cup C_{H,\gamma_H(s)}(\mathbf{P}^+)$ for some H, s; and

(iii) a map $\phi : [2n] \to \coprod_{H \in C} H\backslash G$ by setting

$$\phi(k) = \begin{cases} \phi^{P^+}(k) & \text{if } \; k \leq n \\ \phi^{P^-}(k) & \text{if } \; k > n \text{ and } k-n \notin \cup_{H,s} C_{H,s}(\mathbf{P}^-) \\ \omega^H_{\gamma_H(s)} \phi^{P^-}(k-n) & \text{if } \; k > n \text{ and } k-n \in C_{H,s}(\mathbf{P}^-) \end{cases} .$$

The data (i)–(iii) above satisfy the conditions (b) (i)′ and (iii)′ of Proposition 11 (wit $2n$ instead of the n of the proposition) and therefore determine a unique $P \in R^G(X_{2n})$. It i easy to see that $P \in I(\overline{n})$. Set $\eta((\mathbf{P}^+, \rho(P), \mathbf{P}^-)) = P$.

The proof of the lemma is completed by verifying that the maps ζ and η are inverse t one another. □

In view of the above lemma, we shall feel free, in the sequel, to think of elements c $S(\overline{n}) \times G(\overline{n}) \times S(\overline{n})$ as elements of $I(\overline{n})$, and vice versa.

3. The structure of $A^G_n(d)$

We come now to the representation theory of $A^G_n(d)$.

PROPOSITION 18

Fix $0 \leq k \leq n|G|$ and $\overline{n} \in N_k$. Let $V(\overline{n})$ denote the $\mathbb{C}$-vector space with $S(\overline{n}) \times G(\overline{n})$ c basis.

(a) *The following prescription uniquely defines a representation $\pi_{(\overline{n})}$ of $A^G_n(d)$ on $V(\overline{n}$ temporarily fix an element $\mathbf{S}_0 \in S(\overline{n})$; let $P \in R^G(X_{2n})$, and $(\mathbf{S}, \sigma) \in S(\overline{n}) \times G(\overline{n}$ and suppose $P \cdot (\mathbf{S}, \sigma, \mathbf{S}_0) = d^l Q$ in the algebra $A^G_n(d)$; consider two cases now:*

(i) *if $t(Q) = k$, then $Q \in I(\overline{n})$ and $Q = (\mathbf{S}_1, \sigma_1, \mathbf{S}_0)$ for a unique pair $(\mathbf{S}_1, \sigma_1)$ $S(\overline{n}) \times G(\overline{n})$; in this case, define $\pi_{(\overline{n})}(P)(\mathbf{S}, \sigma) = d^l(\mathbf{S}_1, \sigma_1)$;*

(ii) *if $Q \in I^G_{k-1}$, define $\pi_{(\overline{n})}(P)(\mathbf{S}, \sigma) = 0$.*

(b) *Let $P \in I(\overline{n})$ and $(\mathbf{S}, \sigma) \in S(\overline{n}) \times G(\overline{n})$. Suppose $P = (\mathbf{P}^+, \rho, \mathbf{P}^-)$. Then,*

$$\pi_{(\overline{n})}(P)(\mathbf{S}, \sigma) = D(\mathbf{P}^-, \mathbf{S})(\mathbf{P}^+, \rho\beta^{\mathbf{P}^-}_{\mathbf{S}}\sigma), \qquad (3.$$

where the quantities $D(\mathbf{P}^-, \mathbf{S})$ and $\beta^{\mathbf{P}^-}_{\mathbf{S}}$ are most easily defined by considering tw cases:

Case (i): *For each $H \in C$ such that $\overline{n}(H) \neq 0$, there exist distinct $(R^{P^-} \vee R^S =) R^{P^- \vee}$ equivalence classes, say $C_{H,s}, 1 \leq s \leq \overline{n}(H)$, and a (necessarily unique) permutatic $\gamma_H \in \Sigma_{\overline{n}(H)}$ such that $C_{H,s}(\mathbf{S}) \cup C_{H,\gamma_H(s)}(\mathbf{P}^-) \subset C_{H,s}$ for each $1 \leq s \leq \overline{n}(H)$.*

In this case, define $D(\mathbf{P}^-, \mathbf{S}) = d^{||P^- \vee S|| - k}$, while $\beta^{\mathbf{P}^-}_{\mathbf{S}}$ is defined by the equation

$$(\mathbf{P}^-, 1, \mathbf{P}^-) \cdot (\mathbf{S}, 1, \mathbf{S}) = D(\mathbf{P}^-, \mathbf{S})(\mathbf{P}^-, \beta^{\mathbf{P}^-}_{\mathbf{S}}, \mathbf{S}). \qquad (3.$$

Case (ii): *Suppose the conditions of Case* (i) *are not satisfied.*
In this case, define $D(\mathbf{P}^-, \mathbf{S}) = 0$ and $\beta^{\mathbf{P}^-}_{\mathbf{S}} = 1$.

Proof. (a) We only need to verify that $\pi_{(\overline{n})}(P_1 \cdot P_2) = \pi_{(\overline{n})}(P_1)\pi_{(\overline{n})}(P_2)$ for all P_1, P_2 $R^G(X_{2n})$. Suppose that $(\mathbf{S}, \sigma) \in S(\overline{n}) \times G(\overline{n})$, and $(P_1 \cdot P_2) \cdot (\mathbf{S}, \sigma, \mathbf{S}_0) = d^l Q$. Suppo $P_2 \cdot (\mathbf{S}, \sigma, \mathbf{S}_0) = d^{l_2} Q_2$.

First suppose $t(Q) = k$. It follows that also $t(Q_2) = k$. Deduce now from lemma 7(b) that $Q_2 = (\mathbf{S}_2, \sigma_2, \mathbf{S}_0)$ for some $(\mathbf{S}_2, \sigma_2)$, and that $Q_2 \in I(\overline{n})$. It is also seen – from the associativity of multiplication in $A_n^G(d)$ – that $P_1 \cdot Q_2 = d^{l-l_2}Q$; deduce, as before, that $Q \in I(\overline{n})$ and that $Q = (\mathbf{S}_1, \sigma_1, \mathbf{S}_0)$ for some $(\mathbf{S}_1, \sigma_1)$. Hence, we see that $\pi_{(\overline{n})}(P_1 \cdot P_2)(\mathbf{S}, \sigma) = d^l(\mathbf{S}_1, \sigma_1)$, while

$$\begin{aligned}\pi_{(\overline{n})}(P_1)\pi_{(\overline{n})}(P_2)(\mathbf{S}, \sigma) &= d^{l_2}\pi_{(\overline{n})}(P_1)(\mathbf{S}_2, \sigma_2)\\ &= d^l(\mathbf{S}_1, \sigma_1)\\ &= \pi_{(\overline{n})}(P_1 \cdot P_2)(\mathbf{S}, \sigma),\end{aligned}$$

as desired.

Next, suppose $Q \in I_{k-1}^G$, so that $\pi_{(\overline{n})}(P_1 \cdot P_2)(\mathbf{S}, \sigma) = 0$; then it must be the case that either (i) $Q_2 \in I_{k-1}^G$ or (ii) $Q_2 \in I(\overline{n})$, $Q_2 = (\mathbf{S}_2, \sigma_2, \mathbf{S}_0)$ for some $(\mathbf{S}_2, \sigma_2)$, and $P_1 \cdot (\mathbf{S}_2, \sigma_2, \mathbf{S}_0) \in I_{k-1}^G$. In either case, we have $\pi_{(\overline{n})}(P_1)\pi_{(\overline{n})}(P_2)(\mathbf{S}, \sigma) = 0$.

(b) If $P \in I(\overline{n})$ and $(\mathbf{S}, \sigma) \in S(\overline{n}) \times G(\overline{n})$, it is not hard to see that the following conditions are equivalent:

(α) The conditions of case (i) of (b) are satisfied;
(β) If $P \cdot (\mathbf{S}, \sigma, \mathbf{S}_0) = d^lQ$ in $A_n^G(d)$, then $t(Q) = k$;
(γ) $D(\mathbf{P}^-, \mathbf{S}) \neq 0$.

It is clearly enough to prove that eq. (3.7) is satisfied when the three equivalent conditions above are satisfied. If $P \in I(\overline{n})$ and $(\mathbf{S}, \sigma) \in S(\overline{n}) \times G(\overline{n})$, we thus need to verify (under the stated assumptions above) that

$$(\mathbf{P}^+, \rho, \mathbf{P}^-) \cdot (\mathbf{S}, \sigma, \mathbf{S}_0) = D(\mathbf{P}^-, \mathbf{S})(\mathbf{P}^+, \rho\beta_{\mathbf{S}}^{\mathbf{P}^-}\sigma, \mathbf{S}_0),$$

which we shall do, by considering several special cases.

Case 1: $\mathbf{P}^- = \mathbf{S}$ and $\sigma = 1$.

It is seen from the definition of the product in $A_n^G(d)$ that

$$(\mathbf{P}^+, \rho, \mathbf{S}) \cdot (\mathbf{S}, 1, \mathbf{S}_0) = D(\mathbf{S}, \mathbf{S})(\mathbf{P}^+, \rho, \mathbf{S}_0), \tag{3.9}$$

and eq. (3.7) is satisfied in this case, since (3.8) and the same reasoning, that goes in to justify (3.9), shows that $\beta_{\mathbf{S}}^{\mathbf{S}} = 1$.

We note for future reference that, in the same way, we obtain, for arbitrary $\mathbf{S}_1, \mathbf{S}_2 \in S(\overline{n})$ and $\sigma \in G(\overline{n})$:

$$\begin{aligned}(\mathbf{S}_1, \sigma, \mathbf{S}_2) &= \frac{1}{D(\mathbf{S}_1, \mathbf{S}_1)}(\mathbf{S}_1, 1, \mathbf{S}_1) \cdot (\mathbf{S}_1, \sigma, \mathbf{S}_2)\\ &= \frac{1}{D(\mathbf{S}_2, \mathbf{S}_2)}(\mathbf{S}_1, \sigma, \mathbf{S}_2) \cdot (\mathbf{S}_2, 1, \mathbf{S}_2).\end{aligned} \tag{3.10}$$

Case 2: $\mathbf{P}^- = \mathbf{S}$ and σ is arbitrary.

Thus, we have to verify that

$$(\mathbf{P}^+, \rho, \mathbf{P}^-) \cdot (\mathbf{P}^-, \sigma, \mathbf{S}_0) = D(\mathbf{P}^-, \mathbf{P}^-)(\mathbf{P}^+, \rho\sigma, \mathbf{S}_0), \tag{3.11}$$

and this is really the heart of the computation.

Let us write $P_1 = (\mathbf{P}^-, \sigma, \mathbf{S}_0)$ and $P \cdot P_1 = D(\mathbf{P}^-, \mathbf{P}^-)Q$. Since we are assuming that the conditions $(\alpha) - (\gamma)$ are satisfied, we know from (a) that $Q \in I(\overline{n})$. Suppose

$Q = (\mathbf{Q}^+, \phi, \mathbf{Q}^-)$. We know from Lemma 7 ((a) and (b)) that $\mathbf{Q}^+ = \mathbf{P}^+, \mathbf{Q}^- = \mathbf{S}_0$. Thus we only need to show that $\phi = \rho\sigma$.

Suppose $\rho = ((\rho_H))_{H \in \mathcal{C}}$, where $\rho_H = ((\omega_1^H, \ldots, \omega_{\overline{n}(H)}^H), \gamma_H)$; and that similarly, $\sigma =$ $((\sigma_H))_{H \in \mathcal{C}}$, where $\sigma_H = ((\nu_1^H, \ldots, \nu_{\overline{n}(H)}^H), \kappa_H)$: thus, for each $H \in \mathcal{C}$, we hav $\omega_s^H, \nu_s^H \in N(H)/H, 1 \leq s \leq \overline{n}(H)$, and $\gamma_H, \kappa_H \in \Sigma_{\overline{n}(H)}$.

The construction in the proof of Proposition 17, when unravelled, says that the grou element ρ is related to the relation $P \in R^G(X_{2n})$ by the following requirement, and that is determined by this requirement:

For all $H \in \mathcal{C}, 1 \leq s \leq \overline{n}(H)$, we have:

$$[(\sigma^{P^+}_{C_{H,s}(P^+)}, i(C_{H,s}(P^+)))]_P \supset (\omega_s^H \sigma^{P^-}_{C_{H,\gamma_H^{-1}(s)}(P^-)} \times \{i(C_{H,\gamma_H^{-1}(s)}(P^-))\}).$$

Similarly, we see that for all $H \in \mathcal{C}, 1 \leq s \leq \overline{n}(H)$:

$$[(\sigma^{P^-}_{C_{H,t}(P^-)}, i(C_{H,t}(P^-)))]_{P_1} \supset (\nu_t^H \sigma^{S_0}_{C_{H,\kappa_H^{-1}(t)}(S_0)} \times \{i(C_{H,\kappa_H^{-1}(t)}(S_0))\}).$$

Now, set $t = \gamma_H^{-1}(s)$ in the last inclusion, and use the G-invariance of the relation P_1 t deduce that for all H and s, we have:

$$\begin{aligned}&[(\omega_s^H \sigma^{P^-}_{C_{H,\gamma_H^{-1}(s)}(P^-)} \times i(C_{H,\gamma_H^{-1}(s)}(P^-)))]_{P_1}\\ &\quad \supseteq (\omega_s^H \nu^H_{\gamma_H^{-1}(s)} \sigma^{S_0}_{C_{H,\kappa_H^{-1}(\gamma_H^{-1}(s))}(S_0)} \times \{i(C_{H,\kappa_H^{-1}(\gamma_H^{-1}(s))}(S_0))\}).\end{aligned}$$

Hence, we see that for all H, s, we have:

$$\begin{aligned}&[(\sigma^{P^+}_{C_{H,s}(P^+)}, i(C_{H,s}(P^+)))]_Q\\ &\quad \supset (\omega_s^H \nu^H_{\gamma_H^{-1}(s)} \sigma^{Q^-}_{C_{H,\kappa_H^{-1}(\gamma_H^{-1}(s))}(Q^-)} \times \{i(C_{H,(\gamma_H\kappa_H)^{-1}(s)}(Q^-))\}).\end{aligned}$$

Since this property determines the group element ϕ, we see that $\phi = ((\phi_H))$, wit $\phi_H = ((\chi_1^H, \ldots, \chi_{\overline{n}(H)}^H), \lambda_H)$, where $\chi_s^H = \omega_s^H \nu^H_{\gamma_H^{-1}(s)}$ and $\lambda_H = \gamma_H \kappa_H$; in other word $\phi_H = \rho_H \sigma_H$, the product being computed in the semi-direct product. (This is the reaso for introducing the semi-direct products.)

Case 3: $\mathbf{P}^-, \rho, \mathbf{S}, \sigma$ arbitrary.

Compute as follows:

$$\begin{aligned}(\mathbf{P}^+, \rho, \mathbf{P}^-) \cdot (\mathbf{S}, \sigma, \mathbf{S}_0) &= \frac{(\mathbf{P}^+, \rho, \mathbf{P}^-) \cdot (\mathbf{P}^-, 1, \mathbf{P}^-) \cdot (\mathbf{S}, 1, \mathbf{S}) \cdot (\mathbf{S}, \sigma, \mathbf{S}_0)}{D(\mathbf{P}^-, \mathbf{P}^-)D(\mathbf{S}, \mathbf{S})}\\ &= \frac{D(\mathbf{P}^-, \mathbf{S})(\mathbf{P}^+, \rho, \mathbf{P}^-) \cdot (\mathbf{P}^-, \beta_{\mathbf{S}}^{\mathbf{P}^-}, \mathbf{S}) \cdot (\mathbf{S}, \sigma, \mathbf{S}_0)}{D(\mathbf{P}^-, \mathbf{P}^-)D(\mathbf{S}, \mathbf{S})}\\ &= D(\mathbf{P}^-, \mathbf{S})(\mathbf{P}^+, \rho\beta_{\mathbf{S}}^{\mathbf{P}^-}\sigma, \mathbf{S}_0),\end{aligned}$$

where we have used both the equations (3.10) in the first step, the definition of β (s equation (3.8)) in the second step, and equation (3.11) twice in the last step. [

The next lemma is needed to ensure that that the algebra $A_n^G(d)$ is semisimple at all b a finite number of values of d.

Lemma 19. *Let $C = ((c^i_j))$ be a square matrix and suppose $c^i_j = d^{n^i_j}$, where d is a complex parameter, and the matrix $((n^i_j))$ satisfies the following conditions*:

(i) $n^i_j \in \{-\infty, 0, 1, 2, \ldots\}$,
(ii) $n^i_i \geq \max\{0, n^i_j\}$ $\forall i, j$; *and*
(iii) *if* $i \neq j$ *and* $n^i_j = n^i_i$, *then* $n^i_i < n^j_j$.

Then $\det C$ *is a monic polynomial in d; in particular, the matrix C is non-singular when we substitute all but finitely many possible complex numbers for the parameter d.*

Proof. We shall show that the monomial in d obtained as the 'diagonal product' of C corresponding to any permutation σ which is distinct from the identity permutation, has degree strictly smaller than the degree of the 'main diagonal product' (which corresponds to the identity permutation).

Since any such σ is expressible as a product of disjoint cycles, and since we have assumed that $n^i_i \geq 0$ (so that there is no problem of multiplying by 0), it is enough to (consider the case when σ is just a cycle, and) prove that if $i, j, k, \ldots, r, s$ is a collection of (two or more) distinct indices, then

$$(n^i_j + n^j_k + \cdots + n^r_s + n^s_i) < (n^i_i + n^j_j + \cdots + n^r_r + n^s_s). \tag{3.12}$$

However, we have termwise inequalities:

$$n^i_j \leq n^i_i,\ n^j_k \leq n^j_j, \ldots, n^r_s \leq n^r_r,\ n^s_i \leq n^s_s. \tag{3.13}$$

Since the hypothesis (ii) guarantees that the right side of (3.12) is a finite quantity (i.e., not equal to $-\infty$), the only way that the inequality (3.12) can fail to hold is that each of the inequalities in (3.13) is actually an equality; in that case, the assumption (iii) will imply that $n^i_i < n^j_j < \cdots < n^r_r < n^s_s < n^i_i$. This contradiction completes the proof of the lemma. □

PROPOSITION 20

Let $\overline{n} \in N_k$. *The equation* $(\Gamma(\tau))(\mathbf{R}, \rho) = (\mathbf{R}, \rho\tau^{-1})$ *defines a representation* Γ *of* $G(\overline{n})$ *on* $V(\overline{n})$. *Let* $\pi_{(\overline{n})}$ *be the representation of* $A^G_n(d)$ *described in Proposition* 18. *Then,*

(i) $\pi_{(\overline{n})}(A^G_n(d)) \subset \Gamma(G(\overline{n}))'$.
(ii) *Consider the matrix C with rows and columns indexed by* $S(\overline{n}) \times G(\overline{n})$, *defined – using the notation of Proposition* 18*(b) – by*

$$C((\mathbf{R}, \rho), (\mathbf{S}, \sigma)) = \delta_{\sigma, \rho\beta^{\mathbf{R}}_{\mathbf{S}}} D(\mathbf{R}, \mathbf{S}). \tag{3.14}$$

Then the matrix C satisfies the hypothesis of Lemma 19; *and if d is such that the matrix C is invertible, then*

$$\pi_{(\overline{n})}(A^G_n(d)) = \pi_{(\overline{n})}(\text{span } I(\overline{n})) = \Gamma(G(\overline{n}))'.$$

Proof. (i) Note that $\Gamma(\tau)(\mathbf{R}, \rho) = 1/D(\mathbf{S}_0, \mathbf{S}_0)\pi_{(\overline{n})}(\mathbf{R}, \rho, \mathbf{S}_0)(\mathbf{S}_0, \tau^{-1})$, for each $\tau \in G(\overline{n})$, and $(\mathbf{R}, \rho) \in S(\overline{n}) \times G(\overline{n})$; assertion (i) of the proposition is a consequence of the fact that 'left multiplication' commutes with 'right multiplication'. (ii) It is clear that $C((\mathbf{R}, \rho), (\mathbf{S}, \sigma)) = d^{N((\mathbf{R}, \rho), (\mathbf{S}, \sigma))}$, where N is the matrix defined by

$$N((\mathbf{R}, \rho), (\mathbf{S}, \sigma)) = \begin{cases} ||R \vee S|| - k & \text{if } D(\mathbf{R}, \mathbf{S}) \neq 0 \text{ and } \sigma = \rho\beta^{\mathbf{R}}_{\mathbf{S}} \\ -\infty & \text{otherwise} \end{cases}.$$

Notice first that $\beta^{\mathbf{R}}_{\mathbf{R}} = 1$, and that consequently,

$$N((\mathbf{R},\rho),(\mathbf{R},\rho)) = ||R|| - k \geq \max\{0, N((\mathbf{R},\rho),(\mathbf{S},\sigma))\}\ \forall(\mathbf{R},\rho),(\mathbf{S},\sigma);$$

thus N satisfies conditions (i) and (ii) of lemma 19.

Next, suppose $N((\mathbf{R},\rho),(\mathbf{R},\rho)) = N((\mathbf{R},\rho),(\mathbf{S},\sigma))$ for some $(\mathbf{R},\rho) \neq (\mathbf{S},\sigma)$. In particular, this means that the right side is not equal to $-\infty$, and hence, $D(\mathbf{R},\mathbf{S}) \neq 0$, $\sigma = \rho\beta^{\mathbf{R}}_{\mathbf{S}}$, and $||R \vee S|| = ||R||$. It follows that $R \vee S = R$, i.e., $S \leq R$.

Suppose, if possible, that $R = S$. The condition $D(\mathbf{R},\mathbf{S}) \neq 0$ is then seen to imply that $\mathbf{R} = \mathbf{S}$; then the condition $\sigma = \rho\beta^{\mathbf{R}}_{\mathbf{S}}$ is seen to imply (since $\beta^{\mathbf{R}}_{\mathbf{R}} = 1$) that $\sigma = \rho$; in other words, $(\mathbf{R},\rho) = (\mathbf{S},\sigma)$, contradicting the hypothesis; hence, indeed $R \neq S$.

Then it follows from Lemma 1(b) that $||R|| < ||S||$, and hence that

$$N((\mathbf{R},\rho),(\mathbf{R},\rho)) = ||R|| - t < ||S|| - t = N((\mathbf{S},\sigma),(\mathbf{S},\sigma)),$$

thereby completing the verification that C satisfies the conditions of lemma 19.

So, we assume, in the rest of this proof, that $d \in \mathbb{C}$ is such that the matrix C is invertible. We shall, in what follows, identify a linear operator, say T, on $V(\overline{n})$, with its matrix $((T^{(\mathbf{R},\rho)}_{(\mathbf{S},\sigma)}))$ with respect to the basis $S(\overline{n}) \times G(\overline{n})$. (Thus, $T(\mathbf{S},\sigma) = \sum_{(\mathbf{R},\rho)} T^{(\mathbf{R},\rho)}_{(\mathbf{S},\sigma)}(\mathbf{R},\rho)$.)

Now, the matrix of a typical element of $\Gamma(G(\overline{n}))'$ has the form

$$X((\mathbf{R},\rho),(\mathbf{S},\sigma)) = x^{(\rho\sigma^{-1})}(\mathbf{R},\mathbf{S}),$$

where $\{x^{(\tau)} : \tau \in G(\overline{n})\}$ is a collection of arbitrary matrices with rows and columns indexed by $S(\overline{n})$.

Hence, in order to prove (ii), it will suffice to prove that given an arbitrary collection $\{x^{(\tau)} : \tau \in G(\overline{n})\}$ of matrices with rows and columns indexed by $S(\overline{n})$, then there exist complex scalars $a(\mathbf{Q},\rho,\mathbf{R}), \mathbf{Q},\mathbf{R} \in S(\overline{n}), \rho \in G(\overline{n})$ such that

$$\left(\sum_{\mathbf{Q},\rho,\mathbf{R}} a(\mathbf{Q},\rho,\mathbf{R})\pi_{(\overline{n})}(\mathbf{Q},\rho,\mathbf{R})\right)((\mathbf{S}_1,\sigma_1),(\mathbf{S},\sigma)) = x^{(\sigma_1\sigma^{-1})}(\mathbf{S}_1,\mathbf{S}), \tag{3.15}$$

for all $(\mathbf{S}_1,\sigma_1),(\mathbf{S},\sigma) \in S(\overline{n}) \times G(\overline{n})$.

Fix $\mathbf{S}_1 \in S(\overline{n})$, and define $y^{(\mathbf{S}_1)}(\mathbf{S},\tau) = x^{(\tau)}(\mathbf{S}_1,\mathbf{S})$; due to the assumed invertibility of the matrix C, there exists a unique collection $\{z^{(\mathbf{S}_1)}(\mathbf{R},\rho) : (\mathbf{R},\rho) \in S(\overline{n}) \times G(\overline{n})\}$ of complex numbers such that

$$\sum_{\mathbf{R},\rho} z^{(\mathbf{S}_1)}(\mathbf{R},\rho)C((\mathbf{R},\rho),(\mathbf{S},\tau)) = y^{(\mathbf{S}_1)}(\mathbf{S},\tau), \tag{3.16}$$

for all $\mathbf{S}_1,\mathbf{S},\tau$.

Also note, from (3.7) and the definition of C, that

$$\pi_{(\overline{n})}((\mathbf{Q},\rho,\mathbf{R}))((\mathbf{S}_1,\sigma_1),(\mathbf{S},\sigma)) = \delta_{\mathbf{Q},\mathbf{S}_1}C((\mathbf{R},\rho),(\mathbf{S},\sigma_1\sigma^{-1})).$$

Now set $a(\mathbf{S}_1,\rho,\mathbf{R}) = z^{(\mathbf{S}_1)}(\mathbf{R},\rho)$, and compute as follows:

$$\left(\sum_{\mathbf{Q},\rho,\mathbf{R}} a(\mathbf{Q},\rho,\mathbf{R})\pi_{(\overline{n})}(\mathbf{Q},\rho,\mathbf{R})\right)((\mathbf{S}_1,\sigma_1),(\mathbf{S},\sigma))$$
$$= \sum_{\mathbf{Q},\rho,\mathbf{R}} a(\mathbf{Q},\rho,\mathbf{R})\delta_{\mathbf{Q},\mathbf{S}_1}C((\mathbf{R},\rho),(\mathbf{S},\sigma_1\sigma^{-1}))$$

$$
\begin{aligned}
&= \sum_{\rho,\mathbf{R}} a(\mathbf{S}_1,\rho,\mathbf{R})C((\mathbf{R},\rho),(\mathbf{S},\sigma_1\sigma^{-1}))\\
&= \sum_{\rho,\mathbf{R}} z^{(\mathbf{S}_1)}(\mathbf{R},\rho)C((\mathbf{R},\rho),(\mathbf{S},\sigma_1\sigma^{-1}))\\
&= y^{(\mathbf{S}_1)}(\mathbf{S},\sigma_1\sigma^{-1})\\
&= x^{(\sigma_1\sigma^{-1})}(\mathbf{S}_1,\mathbf{S})
\end{aligned}
$$

and the proof is complete. □

The matrix that we called C in Proposition 20 really depends on $n,\overline{n}$ and d, and we shall write $C^n_{(\overline{n})}(d)$ (rather than merely C) when we wish to emphasize this dependence in the following; likewise, we shall, when desired, write $\Gamma^n_{(\overline{n})}$ for the representation of $G(\overline{n})$ that we called Γ in Proposition 20.

Theorem 21. *Suppose $d \in \mathbb{C}$ is such that $C^n_{(\overline{n})}(d)$ is invertible, for each $\overline{n} \in N_k, 0 \leq k \leq n|G|$. Then*

$$A^G_n(d) \cong \oplus_{\overline{n}\in N_{[n]}} \oplus_{\pi\in\widehat{G(\overline{n})}} (M_{d_\pi}(\mathbb{C}) \otimes M_{|S(\overline{n})|}(\mathbb{C})). \tag{3.17}$$

In particular, the algebras $A^G_n(d)$ are 'generically' semisimple.

Proof. Let us write $L^n_{(\overline{n})} = \Gamma^n_{(\overline{n})}(G(\overline{n}))'$; then, by Proposition 20 (ii), we have, for all $k,\overline{n}$,

$$\pi_{(\overline{n})}(A^G_n(d)) = \pi_{(\overline{n})}(\text{span } I(\overline{n})) = L^n_{(\overline{n})};$$

further, it is clear from the definition that the representation $\Gamma^n_{(\overline{n})}$ is equivalent to $R_{(\overline{n})} \otimes id_{\mathbb{C}^{|S(\overline{n})|}}$, where $R_{(\overline{n})}$ denotes the right regular representation of $G(\overline{n})$; it follows that $L^n_{(\overline{n})} \cong \mathbb{C}[G(\overline{n})] \otimes_{\mathbb{C}} M_{|S(\overline{n})|}(\mathbb{C})$, and hence that $\dim(L^n_{(\overline{n})}) = |G(\overline{n})| \cdot |S(\overline{n})|^2$; on the other hand, we also know that this is the dimension of $I(\overline{n})$ (since $I(\overline{n})$ has a basis indexed by $S(\overline{n}) \times G(\overline{n}) \times S(\overline{n})$), and consequently, we may conclude that $\pi_{(\overline{n})}$ maps (span $I(\overline{n})$) bijectively onto $L^n_{(\overline{n})}$.

Since each $L^n_{(\overline{n})}$ is clearly semisimple, the proposition will be proved once we establish the following isomorphism of $\mathbb{C}$-algebras:

$$\oplus_{\overline{n}\in N_{[n]}}\pi_{(\overline{n})} : A^G_n(d) \cong \oplus_{\overline{n}\in N_{[n]}} L^n_{(\overline{n})}.$$

Now $\dim A^G_n(d) = \dim(\oplus_{\overline{n}} L^n_{(\overline{n})})$, since $\coprod_{\overline{n}} I(\overline{n})$ is a basis of $A^G_n(d)$; so it suffices to prove surjectivity of $\oplus_{\overline{n}}\pi_{(\overline{n})}$.

So suppose $\oplus_{\overline{n}} x_{(\overline{n})} \in \oplus_{\overline{n}} L^n_{(\overline{n})}$; we shall exhibit $\{a_{(\overline{m})} \in (\text{span } I(\overline{m})) : \overline{m} \in N_{[n]}\}$ such that $(\oplus_{\overline{n}}\pi_{(\overline{n})})(\sum_{\overline{m}} a_{(\overline{m})}) = \oplus_{\overline{n}} x_{(\overline{n})}$. Note that $\pi_{(\overline{n})}(I(\overline{m})) = 0$ whenever either (i) $l < k$, or (ii) $l = k$ and $\overline{m} \neq \overline{n}$ – where $\overline{m} \in N_l, \overline{n} \in N_k$; hence the $a_{(\overline{m})}$'s must satisfy

$$\sum_{l\geq k,\overline{m}\in N_l} \pi_{(\overline{n})}(a_{(\overline{m})}) = x_{(\overline{n})} \ \forall\, \overline{n} \in N_k, 0 \leq k \leq n.$$

Since we know that $\pi_{(\overline{n})}$ maps $I(\overline{n})$ onto $L^n_{(\overline{n})}$, we may inductively define the $a_{(\overline{m})}$'s by just requiring that if $\overline{n} \in N_k$, and if $a_{(\overline{m})}$ has been defined for all $\overline{m} \in N_l, l < k$, then

$$\pi_{(\overline{n})}(a_{(\overline{n})}) = x_{(\overline{n})} - \sum_{l>k,\overline{m}\in N_l} \pi_{(\overline{n})}(a_{(\overline{m})}).$$

□

4. The tower $\{A_n^G(d)\colon n = 1, 2, \ldots\}$

Henceforth, we make the blanket assumption that d satisfies the hypothesis of Theorem 21.

It is a consequence of that theorem that – in the notation of that theorem – the irreducible representations of $A_n^G(d)$ are parametrized by the set $\{(\overline{n}, \pi) : \overline{n} \in N_{[n]}, \pi \in \widehat{G(\overline{n})}\}$. For the sake of future computations, we wish to explicitly write out a model for the irreducible representation corresponding to $(\overline{n}, \pi)$. In the sequel, we write $\mathbb{C}S(\overline{n})$ for the $\mathbb{C}$-vector space with basis $S(\overline{n})$, with $S(\overline{n})$ as before.

Remark 22. (i) We wish to note here that although we used a 'reference element' $\mathbf{S}_0$ in defining the representation $\pi_{(n)}$ of Proposition 18, the definition is actually independent of the element $\mathbf{S}_0$ – at least under our blanket assumption that d satisfies the hypothesis of Theorem 21. This is because: (a) it is seen from eq. (3.7) that the definition of $\pi(P)$ is independent of $\mathbf{S}_0$ at least when $P \in I(\overline{n})$; and (b) for a semi-simple algebra, a representation is uniquely determined by its restriction to any ideal which acts 'non-degenerately'.

(ii) Further, as we shall wish to consider $A_n^G(d)$ for varying n, we shall use a subscript n for symbols used so far, to indicate the dependence on n; thus, we shall talk of $V_n(\overline{n})$, $S_n(\overline{n})$, etc.; also, we shall use the notation $N_{n;k}$ for what we have so far denoted by N_k (see Definition 14(c)).

PROPOSITION 23

Fix $\overline{n} \in N_{n;k}, 0 \leq k \leq n|G|, \pi \in \widehat{G(\overline{n})}$. Let V_π denote the vector space on which π represents $G(\overline{n})$, and define $V(\overline{n}, \pi) = \mathbb{C}S(\overline{n}) \otimes V_\pi$.

(a) *Then the following prescription uniquely defines the structure of an $A_n^G(d)$-module on $V(\overline{n}, \pi)$: let $P \in R^G(X_{2n}), \mathbf{S} \in S(\overline{n})$; by the definition of the representation $\pi_{(\overline{n})}$ – see Proposition* 18 *– there exists a unique scalar $C(P, \mathbf{S})$ and an element $(\mathbf{S}_1, \sigma_1) \in S(\overline{n}) \times G(\overline{n})$ such that*

$$\pi_{(\overline{n})}(P)(\mathbf{S}, 1) = C(P, \mathbf{S})(\mathbf{S}_1, \sigma_1),$$

where the 1 *on the left denotes the identity element of $G(\overline{n})$; then let*

$$P \cdot (\mathbf{S} \otimes v) = C(P, \mathbf{S})(\mathbf{S}_1 \otimes \pi(\sigma_1)v).$$

(b) *$V(\overline{n}, \pi)$ is irreducible as a module over the ideal I_k^G (and hence also as an $A_n^G(d)$-module), and further, if $\overline{m} \in N_{n;l}$, then $I(\overline{m})$ acts as* 0, *whenever either* (i) *$l < k$, or* (ii) *$l = k$ and $\overline{m} \neq \overline{n}$.*

(c) *The modules $\{V(\overline{n}, \pi) : 0 \leq k \leq n|G|, \overline{n} \in N_{n;k}, \pi \in \widehat{G(\overline{n})}\}$ are pairwise inequivalent.*

Proof. Suppose $X, Y \in R^G(X_{2n})$ and $\mathbf{S} \in S(\overline{n})$, and suppose that

$$\pi_{(\overline{n})}(Y)(\mathbf{S}, 1) = C(Y, \mathbf{S})(\mathbf{S}_1, \sigma_1); \text{ and}$$
$$\pi_{(\overline{n})}(X)(\mathbf{S}_1, 1) = C(X, \mathbf{S}_1)(\mathbf{S}_2, \sigma_2);$$

it follows that

$$\begin{aligned}\pi_{(\overline{n})}(XY)(\mathbf{S}, 1) &= \pi_{(\overline{n})}(X)\pi_{(\overline{n})}(Y)(\mathbf{S}, 1)\\ &= C(Y, \mathbf{S})\pi_{(\overline{n})}(X)(\mathbf{S}_1, \sigma_1)\end{aligned}$$

$$\begin{aligned} &= C(Y,\mathbf{S})\pi_{(\overline{n})}(X)\Gamma^n_{(\overline{n})}(\sigma_1^{-1})(\mathbf{S}_1,1) \\ &= C(Y,\mathbf{S})\Gamma^n_{(\overline{n})}(\sigma_1^{-1})\pi_{(\overline{n})}(X)(\mathbf{S}_1,1) \\ &= C(Y,\mathbf{S})\Gamma^n_{(\overline{n})}(\sigma_1^{-1})(C(X,\mathbf{S}_1)(\mathbf{S}_2,\sigma_2)) \\ &= C(Y,\mathbf{S})C(X,\mathbf{S}_1)(\mathbf{S}_2,\sigma_2\sigma_1); \end{aligned}$$

it follows from this that $C(XY,\mathbf{S}) = C(Y,\mathbf{S})C(X,\mathbf{S}_1)$.

Now deduce from the definitions that

$$\begin{aligned} Y\cdot(\mathbf{S}\otimes v) &= C(Y,\mathbf{S})(\mathbf{S}_1\otimes\pi(\sigma_1)v); \\ X\cdot(\mathbf{S}_1\otimes w) &= C(X,\mathbf{S}_1)(\mathbf{S}_2\otimes\pi(\sigma_2)w); \end{aligned}$$

and hence that

$$\begin{aligned} XY\cdot(\mathbf{S}\otimes v) &= C(Y,\mathbf{S})C(X,\mathbf{S}_1)(\mathbf{S}_2\otimes\pi(\sigma_2\sigma_1)v) \\ &= C(Y,\mathbf{S})X\cdot(\mathbf{S}_1\otimes\pi(\sigma_1)v) \\ &= X\cdot(C(Y,\mathbf{S})(\mathbf{S}_1\otimes\pi(\sigma_1)v)) \\ &= X\cdot(Y\cdot(\mathbf{S}\otimes v)); \end{aligned}$$

this proves that the 'representation' is multiplicative; the verification of linearity is trivial.

(b) and (c) It is clear that I_{k-1} acts as 0 on $V(\overline{n},\pi)$. Further, if $\overline{m}\in N_{n;k}$, it is a consequence of Lemma 15 that $\pi_{(\overline{n})}(I(\overline{m})) = 0$ for $\overline{m}\neq\overline{n}$, and hence $I(\overline{m})$ also acts as 0 on $V(\overline{n},\pi)$, if $\overline{m}\neq\overline{n}$. On the other hand, $I(\overline{n})$ does not act as 0 on $V(\overline{n},\pi)$, since $\pi_{(\overline{n})}$ is injective on $I(\overline{n})$. It follows from the preceding statements that if $\overline{m}\in N_{n;l}$, then $V(\overline{n},\pi)$ and $V(\overline{m},\chi)$ are inequivalent $A_n^G(d)$-modules, unless $\overline{n}=\overline{m}$.

In order to complete the proof of the proposition, we shall verify – and this is clearly sufficient – that if $T : V(\overline{n},\pi)\to V(\overline{n},\chi)$ is an $I(\overline{n})$-linear map, where $\pi,\chi\in\widehat{G(\overline{n})}$, then

$$T = \begin{cases} \lambda\, id_{V(\overline{n},\pi)} & \text{if } \pi=\chi, \\ 0 & \text{if } \pi \text{ is not equivalent to } \chi, \end{cases}$$

for some $\lambda\in\mathbb{C}$.

Suppose $\{e_j : 1\leq j\leq d_\pi\}$ (resp., $\{f_i : 1\leq i\leq d_\chi\}$) is an orthonormal basis for V_π (resp., V_χ), and suppose

$$T(\mathbf{S}\otimes e_j) = \sum_{\mathbf{S}_1,i} T^{(\mathbf{S}_1,i)}_{(\mathbf{S},j)}(\mathbf{S}_1\otimes f_i).$$

Let $(\mathbf{Q},\rho,\mathbf{R})\in I(\overline{n})$. Computing $T((\mathbf{Q},\rho,\mathbf{R})\cdot(\mathbf{S}\otimes e_j))$ and $(\mathbf{Q},\rho,\mathbf{R})\cdot T(\mathbf{S}\otimes e_j))$, and equating coefficients of $\mathbf{Q_1}\otimes f_l$, we see that

$$D(\mathbf{R},\mathbf{S})\sum_k \pi_j^k(\rho\beta^{\mathbf{R}}_{\mathbf{S}})T^{(\mathbf{Q_1},l)}_{(\mathbf{Q},k)} = \delta^{\mathbf{Q_1}}_{\mathbf{Q}}\sum_{\mathbf{S}_1,i} T^{(\mathbf{S}_1,i)}_{(\mathbf{S},j)}D(\mathbf{R},\mathbf{S}_1)\chi_i^l(\rho\beta^{\mathbf{R}}_{\mathbf{S}_1}), \tag{4.18}$$

for all possible choices of $\mathbf{Q_1},\mathbf{Q},\mathbf{R},\mathbf{S},l,j,\rho$.

If $\mathbf{Q_1}\neq\mathbf{Q}$, set $R=S,\rho=1$ in eq. (4.18) to deduce that

$$T^{(\mathbf{Q_1},l)}_{(\mathbf{Q},j)} = 0\ \forall l,\ j, \tag{4.19}$$

whenever $\mathbf{Q_1}\neq\mathbf{Q}$.

Writing $T_{\mathbf{Q}}$ for the matrix defined by $T_{\mathbf{Q}} = (((T_{\mathbf{Q}})^l_j))$, where $(T_{\mathbf{Q}})^l_j = T^{(\mathbf{Q},l)}_{(\mathbf{Q},j)}$, we next deduce – on setting $\mathbf{R} = \mathbf{S}$, $\mathbf{Q} = \mathbf{Q}_1$ in eq. (4.18) – that

$$T_{\mathbf{Q}}\pi(\rho) = \chi(\rho)T_{\mathbf{S}} \tag{4.20}$$

for all choices of $\rho, \mathbf{Q}, \mathbf{S}$. Set $\rho = 1$ in eq. (4.20) to find that $T_{\mathbf{Q}} = T_{\mathbf{S}} = T_0$ (say), for all $\mathbf{Q}, \mathbf{S}$; deduce next from (4.20) that T_0 intertwines the representations π and χ, thereby completing the proof. □

Since we wish to now look at the inclusion $A^G_n(d) \subset A^G_{n+1}(d)$, it will be necessary to write $S_n(\overline{n})$ for what we called $S(\overline{n})$ till now. Thus,

$$S_n(\overline{n}) = \{\mathbf{S} = (S; (\{C_{H,s}(\mathbf{S}) : H \in C, 1 \leq s \leq \overline{n}(H)\}) \mid S \in R^G(X_n)\}.$$

Given $P \in R^G(X_{2n})$, define $\widetilde{P} \in R^G(X_{2n+2})$ by 'adding on a set of $|G|$-many vertical lines to the right of P'; more pedantically, if $P \in I_n(\overline{n})$ is given by $P = (\mathbf{P}^+, \rho, \mathbf{P}^-)$, with $\mathbf{P}^\pm \in S_n(\overline{n}), \overline{n} \in N_{[n]}$, then $\widetilde{P} = (\widetilde{\mathbf{P}^+}, \widetilde{\rho}, \widetilde{\mathbf{P}^-}) \in S_{n+1}(\overline{m})$, where (a) $\widetilde{P^\pm} = P^\pm \cup \{((g, n+1), (g, n+1)) : g \in G\}$, (b) $\overline{m}(H) = \overline{n}(H) + \delta_{H,\{1\}}$, and

$$C_{H,s}(\widetilde{\mathbf{P}^\pm}) = \begin{cases} \{n+1\} & \text{if } H = \{1\} \text{ and } s = \overline{n}(\{1\}) + 1 \\ C_{H,s}(\mathbf{P}^\pm) & \text{otherwise;} \end{cases}$$

and (c) with the natural identification of $G(\overline{n})$ as a subgroup of $G(\overline{m})$, we have simply $\widetilde{\rho} = \rho$. (Note that (i) G^t sits as the subgroup of G^{t+1} consisting of those elements with last co-ordinate equal to 1, (ii) Σ_t sits as the subgroup of Σ_{t+1} consisting of those permutations which fix $t+1$, (iii) the semi-direct product $G^t \rtimes_s \Sigma_t$ naturally embeds in $G^{t+1} \rtimes \Sigma_{t+1}$ in a manner that is consistent with (i) and (ii) above, and (iv) with $\overline{n}$ and $\overline{m}$ as above, there is a group K such that $G(\overline{n}) = K \rtimes (G^t \rtimes_s \Sigma_t)$ and $G(\overline{m}) = K \times (G^{t+1} \rtimes_s \Sigma_{t+1})$. Later, we shall need the analogous and slightly more general fact that if $\overline{n}, \overline{m} \in N_{[n]}$ and if $\overline{n}(H) \leq \overline{m}(H)$ $\forall H \in C$, then $G(\overline{n})$ may be regarded as a subgroup of $G(\overline{m})$.)

Given $H_0 \in C$, we shall write 1_{H_0} for the function on C which is equal to one at H_0 and 0 elsewhere. In the sequel, we shall specify elements $\mathbf{S} \in S_{n+1}(\overline{m})$ thus: (a) by specifying the data of (i) an element $R^S \in R([n+1])$, (ii) a mapping $[n+1]/R \ni C \mapsto H^S_C \in C$, and (iii) a map ϕ^S defined on $[n+1]$ and taking values in right-cosets of the H^S_C's satisfying the conditions of Proposition 11; and demanding that $S \in R^G(X_{n+1})$ is the unique element corresponding to the data (i)–(iii) as in Proposition 11; and (b) by specifying an explicitly labelled collection $\{C_{H,s}(\mathbf{S}) : 1 \leq s \leq \overline{m}(H), H \in C\}$ of R^S-equivalence classes such that $C_{H,s}(\mathbf{S})$ is assigned to H under the assignment of (a)(ii), and such that the labelling satisfies the condition (2.6).

It will be convenient to have a 'standard' or 'reference' element of each $S_n(\overline{n})$; we specify such an element in the following definition.

DEFINITION 24

Once and for all, fix some total order on the class C. Fix $n \in \{1, 2, \ldots\}$, and $\overline{n} \in N_{[n]}$ Then $\overline{n}$ uniquely specifies distinct elements $H_1, H_2, \ldots, H_l$ of C such that:

(i) $H_1 < H_2 < \cdots < H_l$ (with respect to the chosen total order on C); and
(ii) $\overline{n}(H) \neq 0 \Leftrightarrow H \in \{H_j : 1 \leq j \leq l\}$.

Suppose $\overline{n}(H_j) = \nu_j$; set $\mu_j = \sum_{k=1}^{j} \nu_k$. Then, define $\mathbf{S}_0(n, \overline{n}) \in S_n(\overline{n})$ (which we shal simply abbreviate to $\mathbf{S}_0$ if $n, \overline{n}$ are clear from the context) as follows:

(a) (i) R^{S_0} is the 'identity' equivalence relation on $[n]$, all of whose equivalence classes are singletons;

(ii) $H^{S_0}_{\{k\}} = \begin{cases} H_j & \text{if } \mu_{j-1} < k \leq \mu_j \\ \{1\} & \text{if } \mu_l < k; \end{cases}$

(iii) $\phi^{S_0}(k) = H^{S_0}_{\{k\}} \forall k$; and

(b) $C_{H_j,s}(\mathbf{S}_0) = \{\mu_{j-1} + s\}$, for $1 \leq s \leq \nu_j, 1 \leq j \leq l$.

With a view to decomposing $V_{n+1}(\overline{n}, \chi)$ as an $A_n^G(d)$-module, we shall now proceed to construct several $A_n^G(d)$-linear maps from $V_n(\overline{m}, \pi)$ to $V_{n+1}(\overline{n}, \chi)$, for appropriate $\overline{m}$ and π. The basic idea behind the construction of these intertwiners is the old one that 'right-multiplications commute with left-multiplications'.

Recall, from Definition 9 that for every $H \in C$, we have chosen a fixed set $\{\sigma_\kappa^H : 1 \leq \kappa \leq [G : N(H)]\}$ of coset-representatives for $N(H)\backslash G$.

Lemma 25. *Fix* $\overline{n} \in N_{[n+1]}, H_0 \in C$ *such that* $\overline{n}(H_0) > 0$, *and* $\sigma \in \{\sigma_\kappa^{H_0} : 1 \leq \kappa \leq [G : N(H_0)]\}$.

(1) *If* $\mathbf{Q} \in S_n(\overline{n} - 1_{H_0})$, *define* $\alpha_{H_0,\sigma}(\mathbf{Q}) = \mathbf{S}$, *thus*:

(a) (i) $R^S = R^Q \cup \{(n+1, n+1)\}$,

(ii) $H^S_C = \begin{cases} H^Q_C & \text{if } C \subset [n] \\ H_0 & \text{if } C = \{n+1\} \end{cases}$

(iii) $\phi^S(i) = \begin{cases} \phi^Q(i) & \text{if } i \leq n \\ H_0\sigma & \text{if } i = n+1 \end{cases}$;

(b) $C_{H,s}(\mathbf{S}) = \begin{cases} C_{H,s}(\mathbf{Q}) & \text{if } H \neq H_0 \text{ or } H = H_0, 1 \leq s < \overline{n}(H_0) \\ \{n+1\} & \text{if } H = H_0, s = \overline{n}(H_0) \end{cases}$.

Then $\alpha_{H_0,\sigma}$ *is a* 1–1 *map of* $S_n(\overline{n} - 1_{H_0})$ *into* $S_{n+1}(\overline{n})$.

(2) *Conversely, if* $\mathbf{S} \in S_{n+1}(\overline{n})$, *and if the singleton* $\{n+1\}$ *is an* R^S*-equivalence class which is one of the 'distinguished classes' – meaning that* $\{n+1\} = C_{H_0,s_0}(\mathbf{S})$ *(for a necessarily unique* $H_0 \in C$ *and a unique integer* s_0 *necessarily equal to* $\overline{n}(H_0)$*) – then there exists a unique* $H_0 \in C$ *(namely the one just discussed), a unique* σ, *and a unique* $\mathbf{Q} \in S_n(\overline{n} - 1_{H_0})$ *such that* $\alpha_{H_0,\sigma}(\mathbf{Q}) = \mathbf{S}$.

(3) *Let* $\pi \in \widehat{G(\overline{n} - 1_{H_0})}, \chi \in \widehat{G(\overline{n})}$, *and suppose* $L : V_\pi \to V_\chi$ *is a non-zero* $G(\overline{n} - 1_{H_0})$*-linear operator. (Note that* $G(\overline{n} - 1_{H_0}) \subset G(\overline{n})$, *so the above sentence makes sense.) Then the equation*

$$(A_{H_0,\sigma}(L))(\mathbf{T} \otimes \xi) = \alpha_{H_0,\sigma}(\mathbf{T}) \otimes L\xi \tag{4.21}$$

defines a non-zero $A_n^G(d)$*-linear operator* $A_{H_0,\sigma}(L) : V_n(\overline{n} - 1_{H_0}, \pi) \to V_{n+1}(\overline{n}, \chi)$.

Proof. (1) and (2): It should be clear from the definitions that indeed $\alpha_{H_0,\sigma}$: $S_n(\overline{n} - 1_{H_0}) \to S_{n+1}(\overline{n})$. To complete the proof, we only need to verify injectivity. On the other hand, the statement (2) is also fairly obvious, and explicitly contains the specification of the range of the map $\alpha_{H_0,\sigma}$, as well as the assertion that any point in this range admits a unique pre-image, i.e., that $\alpha_{H_0,\sigma}$ is 1–1.

(3) We shall find the following notation useful: if $\overline{m} \in N_{[m]}$, we shall write $t(\overline{m}) = \sum_{H \in C} \overline{m}(H)[G : H]$; thus, if $P \in I_m(\overline{m})$, then $t(P) = t(\overline{m})$. Also, let $J_m(\overline{m})$ denote

the linear subspace spanned by $(I_m(\overline{m}) \cup \bigcup_{\{\overline{m}':t(\overline{m}')<t(\overline{m})\}} I_m(\overline{m}'))$; it should be clear that $J_m(\overline{m})$ is an ideal in $A_m^G(d)$ which acts non-degenerately on the module $V_m(\overline{m}, \zeta)$.

Since $A_n^G(d)$ is semi-simple, it will suffice to show that $A_{H_0,\sigma}(L)$ is $J_n(\overline{n} - 1_{H_0})$-linear.

First, suppose $P \in I(\overline{m})$ for $\overline{m} \in N_{[n]}$ with $t(\overline{m}) < t(\overline{n} - 1_{H_0})$; then, we shall show that

$$(A_{H_0,\sigma}(L))(P \cdot (\mathbf{T} \otimes \xi)) = 0 = \tilde{P} \cdot (A_{H_0,\sigma}(L))(\mathbf{T} \otimes \xi))\ \forall \mathbf{T} \in S_n(\overline{n} - 1_{H_0}), \xi \in V_\pi.$$

The fact that the left-side of the above equation is zero is a consequence of the fact that $t(P \cdot (\mathbf{T}, 1, \mathbf{S}_0)) \leq t(P) < t(\overline{n} - 1_{H_0})$, and such elements of the algebra act as zero on the module in question.

As for the right side, it suffices to verify that $t(\tilde{P} \cdot (\alpha_{H_0,\sigma}(\mathbf{T}), 1, \mathbf{S}_0)) < t(\overline{n})$; but notice that the first $n|G|$ strands of this product contribute at most $t(P)$ through-classes, while the last $|G|$ strands contribute exactly $[G : H_0]$ through-classes, and the sum of these two terms is, by hypothesis, less than $t(\overline{n})$.

We need now to verify that

$$(A_{H_0,\sigma}(L))(P \cdot (\mathbf{T} \otimes \xi)) = \tilde{P} \cdot (A_{H_0,\sigma}(L))(\mathbf{T} \otimes \xi))\ \forall \mathbf{T} \in S_n(\overline{n} - 1_{H_0}),\ \xi \in V_\pi,$$

whenever $P \in I_n(\overline{n} - 1_{H_0})$. Thus, suppose $P = (\mathbf{Q}, \rho, \mathbf{R})$, where $\mathbf{Q}, \mathbf{R} \in S_n(\overline{n} - 1_{H_0})$ and $\rho \in G(\overline{n} - 1_{H_0})$. Then, if we write $g = \rho\beta_{\mathbf{T}}^{\mathbf{R}}$, we see that since $(\mathbf{Q}, \rho, \mathbf{R}) \cdot (\mathbf{T}, 1, \mathbf{S}_0) = D(\mathbf{R}, \mathbf{T})(\mathbf{Q}, g, \mathbf{S}_0)$, we have, by definition,

$$\begin{aligned}(A_{H_0,\sigma}(L))(P \cdot (\mathbf{T} \otimes \xi)) &= (A_{H_0,\sigma}(L))(D(\mathbf{R}, \mathbf{T})(\mathbf{Q} \otimes \pi(g)\xi)) \\ &= D(\mathbf{R}, \mathbf{T})\alpha_{H_0,\sigma}(\mathbf{Q}) \otimes L\pi(g)\xi. \end{aligned} \tag{4.22}$$

On the other hand, since

$$\tilde{P} \cdot (A_{H_0,\sigma}(L))(\mathbf{T} \otimes \xi)) = \tilde{P} \cdot (\alpha_{H_0,\sigma}(\mathbf{T}) \otimes L\xi)),$$

in order to evaluate the right side of this equation, we will need to first calculate $(\mathbf{Q}, \rho, \mathbf{R})^{\sim} \cdot (\alpha_{H_0,\sigma}(\mathbf{T}), 1, \mathbf{S}_0)$.

To this end, it will be convenient to introduce the following element of $I_{n+1}(\overline{n})$, which we shall denote by $\tilde{\alpha}_{H_0,\sigma}$:

$$\tilde{\alpha}_{H_0,\sigma} = (\alpha_{H_0,\sigma}(\mathbf{S}_0(n, \overline{n} - 1_{H_0})), 1, \mathbf{S}_0(n + 1, \overline{n})).$$

The point is that

$$(\alpha_{H_0,\sigma}(\mathbf{S}), g, \mathbf{S}_0(n + 1, \overline{n})) = (\mathbf{S}, g, \mathbf{S}_0(n, \overline{n} - 1_{H_0}))^{\sim} \cdot \tilde{\alpha}_{H_0,\sigma}, \tag{4.23}$$

for all $\mathbf{S} \in S_n(\overline{n} - 1_{H_0})$ and any $g \in G(\overline{n} - 1_{H_0})$, where the g on the left side of the equation denotes g when thought of as an element of $G(\overline{n})$ (via the natural inclusion $G(\overline{n} - 1_{H_0}) \subset G(\overline{n})$). (Equation (4.23) is verified by looking at the picture represented by the product on the right side, noting that it does belong to $I_{n+1}(\overline{n})$, and checking that its three ingredients are indeed as given by the left side of (4.23).)

Hence,

$$\begin{aligned}(\mathbf{Q}, \rho, \mathbf{R})^{\sim} \cdot (\alpha_{H_0,\sigma}(\mathbf{T}), 1, \mathbf{S}_0) &= (\mathbf{Q}, \rho, \mathbf{R})^{\sim} \cdot (\mathbf{T}, 1, \mathbf{S}_0)^{\sim} \cdot \tilde{\alpha}_{H_0,\sigma} \\ &= D(\mathbf{R}, \mathbf{T})(\mathbf{Q}, g, \mathbf{S}_0)^{\sim} \cdot \tilde{\alpha}_{H_0,\sigma} \\ &= D(\mathbf{R}, \mathbf{T})(\alpha_{H_0,\sigma}(\mathbf{Q}), g, \mathbf{S}_0).\end{aligned}$$

Hence we may deduce that

$$\begin{aligned}(\mathbf{Q},\rho,\mathbf{R})^{\check{}}\cdot(A_{H_0,\sigma}(L))(\mathbf{T}\otimes\xi)) &= (\mathbf{Q},\rho,\mathbf{R})^{\check{}}\cdot(\alpha_{H_0,\sigma}(\mathbf{T})\otimes L\xi))\\ &= D(\mathbf{R},\mathbf{T})(\alpha_{H_0,\sigma}(\mathbf{Q})\otimes\chi(g)L\xi). \qquad (4.24)\end{aligned}$$

Since $g\in G(\overline{n}-1_{H_0})$, we have $\chi(g)L=L\pi(g)$, and the lemma follows from equations (4.22) and (4.24). □

Lemma 26. *Fix $\overline{n}\in N_{[n]}, H_0\in\mathcal{C}$ and $\sigma\in\{\sigma_\kappa^{H_0}:1\le\kappa\le[G:N(H_0)]\}$.*

(1) *If $\mathbf{Q}\in S_n(\overline{n})$, define $\beta_{H_0,\sigma}(\mathbf{Q})=\mathbf{S}$, thus:*

(a) (i) $R^S=R^Q\cup\{(n+1,n+1)\}$;

(ii) $H_C^S=\begin{cases}H_C^Q & \text{if } C\subset[n]\\ H_0 & \text{if } C=\{n+1\}\end{cases}$;

(iii) $\phi^S(i)=\begin{cases}\phi^Q(i) & \text{if } i\le n\\ H_0\sigma & \text{if } i=n+1\end{cases}$;

(b) $C_{H,s}(\mathbf{S})=C_{H,s}(\mathbf{Q})$ for $H\in\mathcal{C}, 1\le s\le\overline{n}(H)$.

Then $\beta_{H_0,\sigma}$ is a 1–1 map of $S_n(\overline{n})$ into $S_{n+1}(\overline{n})$.

(2) *Conversely, if $\mathbf{S}\in S_{n+1}(\overline{n})$, and if the singleton $\{n+1\}$ is an R^S-equivalence class which is not a 'distinguished class' – meaning that $\{n+1\}\notin\{C_{H,s}(\mathbf{S}):H\in\mathcal{C},\ 1\le s\le\overline{n}(H)\}$ – then there exists a unique $H_0\in\mathcal{C}$, a unique σ, and a unique $\mathbf{Q}\in S_n(\overline{n})$ such that $\beta_{H_0,\sigma}(\mathbf{Q})=\mathbf{S}$.*

(3) *Let $\pi\in\widehat{G(\overline{n})}$. Then the equation*

$$B_{H_0,\sigma}(\mathbf{T}\otimes\xi)=\beta_{H_0,\sigma}(\mathbf{T})\otimes\xi \qquad (4.25)$$

defines a non-zero $A_n^G(d)$-linear operator $B_{H_0,\sigma}:V_n(\overline{n},\pi)\to V_{n+1}(\overline{n},\pi)$.

Proof. The proof is almost identical to that of the last lemma, and so we shall say nothing more about the proof except that we would here want to look at the special element $\tilde{\beta}_{H_0,\sigma}\in I_{n+1}(\overline{n})$ defined by

$$\tilde{\beta}_{H_0,\sigma}=(\beta_{H_0,\sigma}(\mathbf{S}_0),1,\mathbf{S}_0),$$

and the crucial identify it satisfies is

$$(\beta_{H_0,\sigma}(\mathbf{S}),g,\mathbf{S}_0)=(\mathbf{S},g,\mathbf{S}_0)^{\check{}}\cdot\tilde{\beta}_{H_0,\sigma}.$$ □

Remark 27. (1) Fix n and $\overline{n}\in N_{[n+1]}$. Consider two cases now:

Case 1: $\overline{n}\in N_{[n]}$. It is a consequence of lemma 25(2) and lemma 26(2) that if $H_i\in\mathcal{C}$, $\sigma_i\in\{\sigma_\kappa^{H_i}:1\le\kappa\le[G:N(H_i)]\}, 1\le i\le 4$, if $\overline{n}(H_1),\overline{n}(H_2)>0$, and if $(H_1,\sigma_1)\ne(H_2,\sigma_2)$ and $(H_3,\sigma_3)\ne(H_4,\sigma_4)$, then the four sets $\alpha_{H_1,\sigma_1}(S_n(\overline{n}-1_{H_1}))$, $\alpha_{H_2,\sigma_2}(S_n(\overline{n}-1_{H_2}))$, $\beta_{H_3,\sigma_3}(S_n(\overline{n})),\beta_{H_4,\sigma_4}(S_n(\overline{n}))$ are pairwise disjoint. In this case, define $W_{n+1}^0(\overline{n})$ to be the subspace of $\mathbb{C}S_{n+1}(\overline{n})$ spanned by $(\cup_{H,\sigma}\alpha_{H,\sigma}(S_n(\overline{n}-1_H)))\cup(\cup_{H,\sigma}\beta_{H,\sigma}(S_n(\overline{n})))$.

Case 2: $\overline{n}\notin N_{[n]}$. Here also, it is true (and follows from lemma 25(2) and lemma 26(2)) that if $H_i\in\mathcal{C},\sigma_i\in\{\sigma_\kappa^{H_i}:1\le\kappa\le[G:N(H_i)]\}, i=1,2,$ if $\overline{n}(H_1),\overline{n}(H_2)>0$, and if

$(H_1, \sigma_1) \neq (H_2, \sigma_2)$, then the sets $\alpha_{H_1,\sigma_1}(S_n(\overline{n} - 1_{H_1}))$ and $\alpha_{H_2,\sigma_2}(S_n(\overline{n} - 1_{H_2}))$ are disjoint. In this case, define $W^0_{n+1}(\overline{n})$ to be the subspace of $\mathbb{C}S_{n+1}(\overline{n})$ spanned by $(\cup_{H,\sigma}\alpha_{H,\sigma}$ $(S_n(\overline{n} - 1_H)))$.

It also follows from lemma 25(2) and lemma 26(2) that if $\mathbf{S} \in S_{n+1}(\overline{n})$, then $\mathbf{S} \in W^0_{n+1}(\overline{n})$ if and only if $\{n+1\}$ is an R^S-equivalence class.

(2) Fix $\overline{n} \in N_{[n+1]}$ as above, and $\chi \in \widehat{G(\overline{n})}$. Also fix $H \in \mathcal{C}$ such that $\overline{n}(H) > 0$, and fix $\sigma \in \{\sigma^H_\kappa : 1 \leq \kappa \leq [G : N(H)]\}$. Consider V_χ as a $G(\overline{n} - 1_H)$-module and decompose into irreducible submodules; specifically, assume that there exist $G(\overline{n} - 1_H)$-linear maps $L_{\pi,j} : V_\pi \rightarrow V_\chi, 1 \leq j \leq m_\pi, \pi \in \widehat{G(\overline{n} - 1_H)}$ such that V_χ is the direct sum of the ranges of all these maps. Then, we wish to observe that:

$$\oplus_{j,\pi} \text{ range } A_{H,\sigma}(L_{\pi,j}) = \mathbb{C}(\alpha_{H,\sigma}(S_n(\overline{n} - 1_H))) \otimes V_\chi;$$

and the two sides of this equation represent an $A^G_n(d)$-submodule of $V_{n+1}(\overline{n}, \chi)$.

Reason: The sum on the left is a direct sum because, for any fixed π, j, the corresponding 'summand' is a subspace of $\mathbb{C}S_{n+1}(\overline{n}) \otimes L_{\pi,j}(V_\pi)$. This direct sum is, by definition included in the space on the right. To prove the reverse inclusion, it is clearly sufficient to verify that any vector of the form $\alpha_{H,\sigma}(\mathbf{Q}) \otimes L_{\pi,j}\xi, \mathbf{Q} \in S_n(\overline{n} - 1_H)), \xi \in V_\pi$ belongs to the left side, but this is just $A_{H,\sigma}(L_{\pi,j})(\mathbf{Q} \otimes \xi)$.

(3) Clearly, for fixed $\overline{n} \in N_{[n]}, \chi \in \widehat{G(\overline{n})}, H \in \mathcal{C}, \sigma \in \{\sigma^H_\kappa : 1 \leq \kappa \leq [G : N(H)]\}$,

$$\text{range } B_{H,\sigma} = \mathbb{C}(\beta_{H,\sigma}(S_n(\overline{n}))) \otimes V_\chi;$$

and the two sides of this equation represent an $A^G_n(d)$-submodule of $V_{n+1}(\overline{n}, \chi)$.

(4) Define $W^0_{n+1}(\overline{n}, \chi) = W^0_{n+1}(\overline{n}) \otimes V_\chi$. It is now a consequence of (1)–(3) above that

$$W^0_{n+1}(\overline{n}, \chi) = (\oplus_{H,\sigma,j,\pi} \text{ range } A_{H,\sigma}(L_{\pi,j})) \oplus (\oplus_{H,\sigma} \text{ range } B_{H,\sigma}),$$

and that $W^0_{n+1}(\overline{n}, \chi)$ is an $A^G_n(d)$-submodule of $V_{n+1}(\overline{n}, \chi)$.

Before discussing further intertwiners, it will be convenient to describe some coset representatives for some subgroups, and also to discuss a certain natural group action; we do so in the following lemma, whose proof we omit since it is an easy verification.

Lemma 28. *Fix* $\overline{m} \in N_{[m]}, H_0 \in \mathcal{C}$.

(1) *For* $1 \leq s \leq \overline{m}(H_0) + 1$, *and* $f \in H_0 \backslash N(H_0)$, *define an element* $\sigma(s, f) \in G(\overline{m} + 1_{H_0}$ *as follows*:

$$\sigma(s, f) = ((\sigma(s, f))_H),$$

where

$$(\sigma(s, f))_H = \begin{cases} 1 & \text{if } H \neq H_0 \\ ((1, \ldots, f, 1, \ldots, 1); \lambda_s) & \text{if } H = H_0 \end{cases}$$

where the f *occurs in the s-th slot, and* λ_s *is the cycle* $(s, s+1, \ldots, \overline{m}(H_0) + 1)$.

Then, $G(\overline{m} + 1_{H_0}) = \coprod_{(s,f)} \sigma(s, f) G(\overline{m})$.

(2) *Consider the set* $\{1, \ldots, \overline{m}(H_0) + 1\} \times (H_0 \backslash N(H_0))$ *and an element* $g \in G(\overline{m} + 1_{H_0}$ *suppose that*

$$g = ((g_H)), \textit{where } g_H = ((\omega^H_1, \ldots, \omega^H_{(m+1_{H_0})(H)}); \kappa_H).$$

Then, the equations

$$g \cdot (s,f) = (s_1, f_1) \Leftrightarrow s_1 = \kappa_{H_0}(s), f_1 = \omega_{s_1}^{H_0} f$$

define a transitive action of $G(\overline{m} + 1_{H_0})$ *on* $\{1, \dots, \overline{m}(H_0) + 1\} \times (H_0 \backslash N(H_0))$. *In fact,* $g \cdot (s,f) = (s_1, f_1) \Leftrightarrow \sigma(s_1, f_1)^{-1} g \sigma(s,f)) \in G(\overline{m})$; *equivalently, this action may be identified with that of* $G(\overline{m} + 1_{H_0})$ *on* $G(\overline{m} + 1_{H_0})/G(\overline{m})$.

Lemma 29. *Fix* $\overline{n} \in N_{[n]}, H_0 \in \mathcal{C}$ *such that* $\overline{n}(H_0) > 0$; *also fix* $1 \leq s_0 \leq \overline{n}(H_0), f \in H_0 \backslash N(H_0)$ *and* $\sigma \in \{\sigma_\kappa^{H_0} : 1 \leq \kappa \leq [G : N(H_0)]\}$.

(1) *If* $\mathbf{Q} \in S_n(\overline{n})$, *define* $\gamma_{H_0,(\sigma,s_0,f)}(\mathbf{Q}) = \mathbf{S}$, *thus*:

(a) (i) *the* R^S*-equivalence classes are* $C = C_{H_0,s_0}(\mathbf{Q}) \cup \{n+1\}$ *and all the* R^Q*-equivalence classes other than* $C_{H_0,s_0}(\mathbf{Q})$;

(ii) $H_C^S = H_{C \cap [n]}^Q$;

(iii) $\phi^S(i) = \begin{cases} \phi^Q(i) & \textit{if } i \leq n \\ f\sigma & \textit{if } i = n+1 \end{cases}.$

(b) *Define* $C_{H_0,s_0}(\mathbf{S})$ *to be the* C *defined in* (1) (a) (i), *and for* $H \in \mathcal{C}, 1 \leq s \leq \overline{n}(H)$, $(H,s) \neq (H_0, s_0)$, *define* $C_{H,s}(\mathbf{S}) = C_{H,s}(\mathbf{Q})$.

Then $\gamma_{H_0,(\sigma,s_0,f)}$ *is a* 1–1 *map of* $S_n(\overline{n})$ *into* $S_{n+1}(\overline{n})$.

(2) *Conversely, if* $\mathbf{S} \in S_{n+1}(\overline{n})$, *and if* $[n+1]_{R^S}$ *is not a singleton set, and if this* R^S*-equivalence class is a 'distinguished class' – meaning that* $[n+1]_{R^S} = C_{H_0,s_0}(\mathbf{S})$ *for a necessarily unique* (H_0, s_0) *– then there exists a unique* $f \in H_0 \backslash N(H_0)$, $\sigma \in \{\sigma_\kappa^{H_0} : 1 \leq \kappa \leq [G : N(H_0)]\}$ *and a* $\mathbf{Q} \in S_n(\overline{n})$ *such that* $\gamma_{H_0,(\sigma,s_0,f)}(\mathbf{Q}) = \mathbf{S}$.

(3) *Suppose* $\pi, \chi \in \widehat{G(\overline{n})}$, *and suppose* $L : V_\pi \to V_\chi$ *is a non-zero* $G(\overline{n} - 1_{H_0})$*-linear map. Also suppose that* $\sigma \in \{\sigma_\kappa^{H_0} : 1 \leq \kappa \leq [G : N(H_0)]\}$. *Then the equation*

$$\begin{aligned}&(C_{H_0,\sigma}(L))(\mathbf{T} \otimes \xi) \\ &\quad = \sum_{(s,f)} (\gamma_{H_0,(\sigma,s,f)}(\mathbf{T}) \otimes \chi(\sigma(s,f)) L \pi(\sigma(s,f)^{-1}) \xi) + W_{n+1}^0(\overline{n}, \chi),\end{aligned}$$

where the sum ranges over $1 \leq s \leq \overline{n}(H_0)$ *and* $f \in H_0 \backslash N(H_0)$, *defines a non-zero* $A_n^G(d)$*-linear map* $C_{H_0,\sigma}(L) : V_n(\overline{n}, \pi) \to (V_{n+1}(\overline{n}, \chi)/W_{n+1}^0(\overline{n}, \chi))$.

Proof. The statements (1) and (2) are established exactly like their counterparts in Lemma 25 after observing that every element of $H_0 \backslash G$ is uniquely expressible as $f\sigma$ for $f \in H_0 \backslash N(H_0)$ and $\sigma \in \{\sigma_\kappa^{H_0} : 1 \leq \kappa \leq [G : N(H_0)]\}$. For (3), again as in Lemma 25, it will suffice to verify that $C_{H_0,\sigma}(L)$ is $J_n(\overline{n})$-linear. Thus we need to verify that

$$(C_{H_0,\sigma}(L))(P \cdot ((\mathbf{T} \otimes \xi)) = \tilde{P} \cdot ((C_{H_0,\sigma}(L))(\mathbf{T} \otimes \xi)) \ \forall \mathbf{T} \in S_n(\overline{n}), \xi \in V_\pi,$$

whenever either (i) $P \in I_n(\overline{m})$, where $\overline{m} \in N_{[n]}, t(\overline{m}) < t(\overline{n})$, or (ii) $P = (\mathbf{Q}, \rho, \mathbf{R}) \in I_n(\overline{n})$.

We first show that in case (i), both sides of the desired equation reduce to zero. Since $t(P \cdot (\mathbf{T}, 1, \mathbf{S}_0)) \leq t(P) < t(\overline{n})$, it is seen that the left side of the above equation is, indeed, zero. To evaluate the right side, we have to examine such products as $\tilde{P} \cdot (\gamma_{H_0,(\sigma,s,f)}(\mathbf{T}), 1, \mathbf{S}_0)$, and it will suffice to show, therefore, that in this case, either such a product has less than $t(\overline{n})$ through classes, or such a product has exactly $t(\overline{n})$ through classes, in which case its 'top' belongs to $W_{n+1}^0(\overline{n})$. In any case, since the second term of the product has exactly $t(\overline{n})$ through classes, the product can have at most $t(\overline{n})$ through classes. So, we may assume without loss of generality that the product has exactly $t(\overline{n})$ through classes. By the last line of Remark 27(1), it suffices therefore to show that if the 'top' of our product is $\mathbf{S}$, and if $\{n+1\}$ is not an R^S-equivalence class, then the product cannot have $t(\overline{n})$ through classes; but this is an easy consequence of the assumption that $t(P) < t(\overline{n})$.

To discuss the second (and less trivial) case, we will again find it convenient to introduce an auxiliary element $\tilde{\gamma}_{H_0,(\sigma,s,f)}$ which enables us to regard the mapping $\gamma_{H_0,(\sigma,s,f)}$ as a sort of right-multiplication. Thus, define $\tilde{\gamma}_{H_0,(\sigma,s,f)} \in I_{n+1}(\overline{n})$ by

$$\tilde{\gamma}_{H_0,(\sigma,s,f)} = (\gamma_{H_0,(\sigma,s,f)}(\mathbf{S}_0), 1, \mathbf{S}_0).$$

We can now state the desired analogue of eq. (4.23), namely:

$$(\mathbf{S}, g, \mathbf{S}_0)^{\sim} \cdot \tilde{\gamma}_{H_0,(\sigma,s,f)} = (\gamma_{H_0,(\sigma,g\cdot(s,f))}(\mathbf{S}), g, \mathbf{S}_0) \tag{4.26}$$

for all $\mathbf{S} \in S_n(\overline{n} - 1_{H_0}), g \in G(\overline{n} - 1_{H_0})$, where the g on the right is g thought of as an element of $G(\overline{n})$ (via the natural inclusion) and $g \cdot (s, f)$ refers to the action of $G(\overline{n})$ as in Lemma 28(2) applied to $\overline{m} = \overline{n} - 1_{H_0}$. As in the case of eq. (4.23), this equation is also verified by looking at the picture represented by the product on the left side, noting that it does belong to $I_{n+1}(\overline{n})$, and checking that its three ingredients are indeed as given by the right side of (4.26).

Suppose now that $P = (\mathbf{Q}, \rho, \mathbf{R}) \in I_n(\overline{n})$. If we let $g = \rho\beta^{\mathbf{R}}_{\mathbf{T}}$, we find that $(\mathbf{Q}, \rho, \mathbf{R}) \cdot (\mathbf{T}, 1, \mathbf{S}_0) = D(\mathbf{R}, \mathbf{T})(\mathbf{Q}, g, \mathbf{S}_0)$, and hence

$$\begin{aligned}
&(C_{H_0,\sigma}(L))(P \cdot (\mathbf{T} \otimes \xi))\\
&\quad = (C_{H_0,\sigma}(L))(D(\mathbf{R}, \mathbf{T})(\mathbf{Q} \otimes \pi(g)\xi))\\
&\quad = D(\mathbf{R}, \mathbf{T}) \sum_{(s_1,f_1)} (\gamma_{H_0,(\sigma,s_1,f_1)}(\mathbf{Q}) \otimes \chi(\sigma(s_1, f_1))L\pi(\sigma(s_1, f_1)^{-1})\pi(g)\xi)\\
&\qquad\qquad + W^0_{n+1}(\overline{n}, \chi);
\end{aligned} \tag{4.27}$$

writing $(s_1, f_1) = g \cdot (s, f)$, and noting that

$$\begin{aligned}
L\pi(\sigma(s_1, f_1)^{-1})\pi(g) &= L\pi(\sigma(s_1, f_1)^{-1}g)\\
&= L\pi(\sigma(s_1, f_1)^{-1}g\sigma(s, f))\pi(\sigma(s, f)^{-1})\\
&= \chi(\sigma(s_1, f_1)^{-1}g\sigma(s, f))L\pi(\sigma(s, f)^{-1})
\end{aligned}$$

(where we have used the fact that $\sigma(s_1, f_1)^{-1}g\sigma(s, f) \in G(\overline{n} - 1_{H_0})$), we see thus that the right side of eq. (4.27) may be rewritten (after a change of variable) as

$$D(\mathbf{R}, \mathbf{T}) \sum_{(s,f)} (\gamma_{H_0,(\sigma,g\cdot(s,f))}(\mathbf{Q}) \otimes \chi(g\sigma(s, f))L\pi(\sigma(s, f)^{-1})\xi) + W^0_{n+1}(\overline{n}, \chi). \tag{4.28}$$

On the other hand, notice (by two applications of eq. (4.26)) that

$$\begin{aligned}
(\mathbf{Q}, \rho, \mathbf{R})^{\sim} \cdot (\gamma_{H_0,(\sigma,s,f)}(\mathbf{T}), 1, \mathbf{S}_0) \cdot &= (\mathbf{Q}, \rho, \mathbf{R})^{\sim} \cdot (\mathbf{T}, 1, \mathbf{S}_0)^{\sim} \cdot \tilde{\gamma}_{H_0,(\sigma,s,f)}\\
&= ((\mathbf{Q}, \rho, \mathbf{R}) \cdot (\mathbf{T}, 1, \mathbf{S}_0))^{\sim} \cdot \tilde{\gamma}_{H_0,(\sigma,s,f)}\\
&= D(\mathbf{R}, \mathbf{T})(\mathbf{Q}, g, S_0)^{\sim} \cdot \tilde{\gamma}_{H_0,(\sigma,s,f)}\\
&= D(\mathbf{R}, \mathbf{T})(\gamma_{H_0,(\sigma,g\cdot(s,f))}(\mathbf{Q}), g, \mathbf{S}_0).
\end{aligned}$$

Hence, we see that

$$\begin{aligned}&\widetilde{P}\cdot(C_{H_0,\sigma}(L))(\mathbf{T}\otimes\xi))\\ &=(\mathbf{Q},\rho,\mathbf{R})^{\tilde{}}\cdot\sum_{(s,f)}(\gamma_{H_0,(\sigma,s,f)}(\mathbf{T})\otimes\chi(\sigma(s,f))L\pi(\sigma(s,f)^{-1})\xi)+W^0_{n+1}(\overline{n},\chi)\\ &=D(\mathbf{R},\mathbf{T})\sum_{(s,f)}(\gamma_{H_0,(\sigma,g\cdot(s,f))}(\mathbf{Q})\otimes\chi(g\sigma(s,f))L\pi(\sigma(s,f)^{-1})\xi)+W^0_{n+1}(\overline{n},\chi),\end{aligned}$$

and the lemma is proved. □

Remark 30. (1) Fix n and $\overline{n}\in N_{[n+1]}$. Consider two cases now:

Case 1: $\overline{n}\in N_{[n]}$. Fix $H\in\mathcal{C}$ such that $\overline{n}(H)>0$ and define $\overline{W}^1_{n+1}(\overline{n};H)$ to be the subspace of $\mathbb{C}S_{n+1}(\overline{n})/W^0_{n+1}(\overline{n})$ spanned by $\{\gamma_{H,(\sigma,s,f)}(\mathbf{T})+W^0_{n+1}(\overline{n})\colon \mathbf{T}\in S_n(\overline{n}),\sigma\in\{\sigma^H_\kappa\colon 1\le\kappa\le[G:N(H)]\},1\le s\le\overline{n}(H),\ f\in H\backslash N(H)\}$. By Lemma 29(2), this spanning set is a basis, which can alternatively be described as $\{\mathbf{S}+W^0_{n+1}(\overline{n})\colon\mathbf{S}\in S_{n+1}(\overline{n}),[n+1]_{R_S}$ is a distinguished class which is not a singleton, and $H^S_{[n+1]}=H\}$.

Also set $\overline{W}^1_{n+1}(\overline{n})=\sum_{H\in\mathcal{C},\overline{n}(H)>0}\overline{W}^1_{n+1}(\overline{n};H)$, and note (again, by Lemma 29(2)) that this sum of subspaces is a direct sum.

Case 2: $\overline{n}\notin N_{[n]}$. In this case, define $\overline{W}^1_{n+1}(\overline{n})=\{0\}\subseteq\mathbb{C}S_{n+1}(\overline{n})/W^0_{n+1}(\overline{n})$.

In either case, set $W^1_{n+1}(\overline{n})$ be the inverse image, in $\mathbb{C}S_{n+1}(\overline{n})$, under the natural quotient map, of $\overline{W}^1_{n+1}(\overline{n})$, and observe that a basis for $W^1_{n+1}(\overline{n})$ is furnished by $\{\mathbf{S}\in S_{n+1}(\overline{n})\colon [n+1]_{R^S}$ is either a singleton or is a distinguished R^S-class$\}$.
(2) We will need to use the following elementary fact about induced representations. Let $\mathcal{G}_0$ be a subgroup of a finite group $\mathcal{G}$ and $\mathcal{G}=\coprod_{k=1}^n g_k\mathcal{G}_0$ with $g_1=1$. For $\chi\in\hat{\mathcal{G}}$ and $\tilde{\chi}=\mathrm{Ind}_{\mathcal{G}_0\uparrow\mathcal{G}}\mathrm{Res}_{\mathcal{G}\downarrow\mathcal{G}_0}(\chi)$, we may regard $V_{\tilde{\chi}}$ as the space $V_\chi\otimes\mathbb{C}(\mathcal{G}/\mathcal{G}_0)$ with $\mathcal{G}$-action defined by $g(v\otimes g_i\mathcal{G}_0)=\chi(h(g,i))(v)\otimes g_{g(i)}\mathcal{G}_0$ where $gg_i=g_{g(i)}h(g,i)$ with $g_{g(i)}\in\{g_1,\dots,g_k\}$ and $h(g,i)\in\mathcal{G}_0$. Furthermore, for $\pi\in\hat{\mathcal{G}}$, there is a natural bijection between $\mathcal{G}_0$-linear maps $L\colon V_\pi\to V_\chi$ and $\mathcal{G}$-linear maps $\tilde{L}\colon V_\pi\to V_{\tilde{\chi}}$ given by $\tilde{L}(\xi)=\sum_{k=1}^n L\pi(g_k^{-1})(\xi)\otimes g_k\mathcal{G}_0$.
(3) Fix $\overline{n}\in N_{[n]}$ and $H\in C$ such that $\overline{n}(H)>0$. Also fix $\chi\in\widehat{G(\overline{n})}$. Let $\tilde{\chi}=\mathrm{Ind}_{G(\overline{n}-1_H)\uparrow G(\overline{n})}\mathrm{Res}_{G(\overline{n})\downarrow G(\overline{n}-1_H)}(\chi)$, and for appropriate $\pi\in G(\overline{n})$, choose non-zero $G(\overline{n})$-linear maps $L_{\pi,j}\colon V_\pi\to V_{\tilde{\chi}}$ so that the ranges of all these maps yield a direct sum decomposition of $V_{\tilde{\chi}}$. Let $L_{\pi,j}\colon V_\pi\to V_\chi$ be the $G(\overline{n}-1_H)$-linear map which is related to $\tilde{L}_{\pi,j}$ as in the above paragraph. We wish now to assert that

$$\oplus_{\sigma,j,\pi}\text{ range } C_{H,\sigma}(L_{\pi,j})=\overline{W}^1_{n+1}(\overline{n};H)\otimes V_\chi;\qquad(4.29)$$

hence the right side represents an $A^G_n(d)$-submodule of $V_{n+1}(\overline{n},\chi)/W^0_{n+1}(\overline{n},\chi)$.

Reason: Identify $V_{n+1}(\overline{n},\chi)/W^0_{n+1}(\overline{n},\chi)$ with $(\mathbb{C}S_{n+1}(\overline{n})/W^0_{n+1}(\overline{n}))\otimes V_\chi$; the definition of $C_{H,\sigma}(L)$ shows that every summand on the left (of eq. (4.29)) is contained in the space on the right. To see that the sum is direct, define $\Phi\colon\overline{W}^1_{n+1}(\overline{n};H)\otimes V_\chi\to\mathrm{Hom}(\mathbb{C}S_n(\overline{n}),\mathbb{C}(N(H)\backslash G)\otimes V_{\tilde{\chi}})$ as follows: an arbitrary element $Z\in\overline{W}^1_{n+1}(\overline{n};H)\otimes V_\chi$ can be expressed uniquely as $Z=\sum_{\mathbf{T}\in S_n(\overline{n})}\sum_{(\sigma,s,f)}\gamma_{H,(\sigma,s,f)}(\mathbf{T})\otimes\xi^{\mathbf{T}}_{(\sigma,s,f)}+W^0_{n+1}(\overline{n},\chi)$, where $\xi^{\mathbf{T}}_{(\sigma,s,f)}\in V_\chi$ for all (σ,s,f); define

$$\Phi(Z)(\mathbf{T})\in\mathbb{C}(N(H)\backslash G)\otimes V_{\tilde{\chi}}=\mathbb{C}(N(H)\backslash G)\otimes V_\chi\otimes\mathbb{C}(G(\overline{n})/G(\overline{n}-1_H))$$

by

$$\Phi(Z)(\mathbf{T}) = \sum_{(\sigma,s,f)} N(H)\sigma \otimes \chi(\sigma(s,f)^{-1})(\xi^{\mathbf{T}}_{(\sigma,s,f)}) \otimes \sigma(s,f)G(\overline{n}-1_H).$$

The map Φ is clearly injective since knowledge of all $\Phi(Z)(\mathbf{T})$ determines all the $\xi^{\mathbf{T}}_{(\sigma,s,f)}$ and hence $\Phi(Z)$ determines Z; i.e., Φ is an injective (clearly linear) map. Further, it is easy to see that if $Z \in$ range $C_{H,\sigma}(L)$ then range $\Phi(Z) \subseteq \mathbb{C}(N(H)\sigma) \otimes$ range $\tilde{L}$. Together with the injectivity of Φ and the choice of $L_{\pi,j}$, this implies that the ranges of the $C_{H,\sigma}(L_{\pi,j})$'s form a direct sum. Finally, a dimension count – using Frobenius reciprocity for the dimension of the left side (of eq. (4.29)), and the explicitly listed basis for the first factor of the tensor product on the right – shows that both sides of eq. (4.29) have dimension $|S_n(\overline{n})|d_\chi[G(\overline{n}) : G(\overline{n}-1_H)][G : N(H)]$; and therefore the direct sum on the left exhausts the space on the right.

(4) Define $\overline{W}^1_{n+1}(\overline{n},\chi) = \overline{W}^1_{n+1}(\overline{n}) \otimes V_\chi$, and as before, let $W^1_{n+1}(\overline{n},\chi)$ be the inverse image, in $V_{n+1}(\overline{n},\chi)$, of $\overline{W}^1_{n+1}(\overline{n},\chi)$, under the natural quotient mapping. If $\overline{n} \in N_{[n]}$ (as in (3) above), it then follows that

$$\overline{W}^1_{n+1}(\overline{n},\chi) = \oplus_{H\in\mathcal{C},\overline{n}(H)>0} \oplus_{\sigma,\pi,j} \text{ range } C_{H,\sigma}(L_{\pi,j});$$

since $\overline{W}^1_{n+1}(\overline{n},\chi) = \{0\}$ if $\overline{n} \notin N_{[n]}$, we find thus, in any case, that $W^1_{n+1}(\overline{n},\chi)$ is an $A^G_n(d)$-submodule of $V_{n+1}(\overline{n},\chi)$.

Lemma 31. *Fix* $\overline{n} \in N_{[n]}, H_0 \in \mathcal{C}, 1 \le s_0 \le \overline{n}(H_0)+1,\ f \in H_0\backslash N(H_0)$ *and* $\sigma \in \{\sigma^{H_0}_\kappa : 1 \le \kappa \le [G : N(H_0)]\}$.

(1) *If* $\mathbf{Q} \in S_n(\overline{n}+1_{H_0})$, *define* $\delta_{H_0,(\sigma,s_0,f)}(\mathbf{Q}) = \mathbf{S}$, *thus*:

(a) (i) R^S *is defined exactly as in Proposition* 29(1) (a) (i);

(ii) $H^S_{\mathcal{C}} = H^Q_{\mathcal{C}\cap[n]}$;

(iii) $\phi^S(i) = \begin{cases} \phi^Q(i) & \text{if } i \le n \\ f\sigma & \text{if } i = n+1 \end{cases}$.

(b) *Define*

$$C_{H,s}(\mathbf{S}) = \begin{cases} C_{H,s}(\mathbf{Q}) & \text{if } H \ne H_0 \text{ or } H = H_0, 1 \le s < s_0 \\ C_{H_0,s+1}(\mathbf{Q}) & \text{if } H = H_0 \text{ and } s_0 \le s \le \overline{n}(H_0). \end{cases}$$

Then $\delta_{H_0,(\sigma,s_0,f)}$ *defines a* 1–1 *map of* $S_n(\overline{n}+1_{H_0})$ *into* $S_{n+1}(\overline{n})$.

(2) *Conversely, if* $\mathbf{S} \in S_{n+1}(\overline{n})$, *and if* $[n+1]_{R^S}$ *is not a singleton set, and if this* R^S-*equivalence class is not a 'distinguished class' – then there exists a unique* $H_0 \in \mathcal{C}$, *a unique* $1 \le s_0 \le \overline{n}(H_0)+1$, *a unique* $f \in H_0\backslash N(H_0)$, *a unique* $\sigma \in \{\sigma^{H_0}_\kappa : 1 \le \kappa \le [G : N(H_0)]\}$ *and a unique* $\mathbf{Q} \in S_n(\overline{n}+1_{H_0})$ *such that* $\delta_{\mathcal{H}_0,(\sigma,s_0,f)}(\mathbf{Q}) = \mathbf{S}$.

(3) *Suppose* $\pi \in G(\overline{n}+1_{H_0}), \chi \in G(\overline{n})$, *and suppose* $L : V_\pi \to V_\chi$ *is a non-zero* $G(\overline{n})$-*linear map. Then the equation*

$$(D_{H_0,\sigma}(L))(\mathbf{T}\otimes\xi) = \sum_{(s,f)} (\delta_{H_0,(\sigma,s,f)}(\mathbf{T}) \otimes L\pi(\sigma(s,f)^{-1})\xi) + W^1_{n+1}(\overline{n},\chi), \tag{4.30}$$

where the sum ranges over all choices $1 \le s \le \overline{n}(H_0), f \in H_0\backslash N(H_0)$, (*with* $\sigma(s,f)$ *as in Lemma 28 applied with* $\overline{m} = \overline{n}+1_{H_0}$), *defines a non-zero* $A^G_n(d)$-*linear map* $D_{H_0,\sigma}(L)$: $V_n(\overline{n}+1_{H_0},\pi) \to (V_{n+1}(\overline{n},\chi)/W^1_{n+1}(\overline{n},\chi))$.

Proof. The statements (1) and (2) are established exactly like their counterparts in Lemma 25. For (3), exactly as in Lemma 25, it will suffice to verify that $D_{H_0,\sigma}(L)$ is $J_n(\overline{n}+1_{H_0})$-linear. Thus we need to verify that

$$(D_{H_0,\sigma}(L))(P\cdot(\mathbf{T}\otimes\xi)) = \tilde{P}\cdot(D_{H_0,\sigma}(L))(\mathbf{T}\otimes\xi))\ \forall\mathbf{T}\in S_n(\overline{n}+1_{H_0}), \xi\in V_\pi,$$

whenever either (i) $P\in I_n(\overline{m})$, where $\overline{m}\in N_{[n]}, t(\overline{m}) < t(\overline{n}+1_{H_0})$, or (ii) $P=(\mathbf{Q},\rho,\mathbf{R})\in I_n(\overline{n}+1_{H_0})$.

We first show that in case (i), both sides of the desired equation reduce to zero. Since $t(P\cdot(\mathbf{T},1,\mathbf{S}_0)) \leq t(P) < t(\overline{n}+1_{H_0})$, it is seen that the left side of the above equation is, indeed, zero. To evaluate the right side, we have to examine such products as $\tilde{P}\cdot(\delta_{H_0,(\sigma,s,f)}(\mathbf{T}),1,\mathbf{S}_0)$, and it will suffice to show, therefore, that in this case, either such a product has less than $t(\overline{n})$ through classes, or such a product has exactly $t(\overline{n})$ through classes, in which case its 'top' belongs to $W^1_{n+1}(\overline{n})$. In any case, since the second term of the product has exactly $t(\overline{n})$ through classes, the product can have at most $t(\overline{n})$ through classes. So, we may assume without loss of generality that the product has exactly $t(\overline{n})$ through classes. By the last line of Remark 30(1), it suffices therefore to show that if the 'top' of our product is $\mathbf{S}$, and if $[n+1]_{R^S}$ is neither a singleton nor a distinguished R^S-class, then the product cannot have $t(\overline{n})$ through classes; but this is an easy consequence of the assumption that $t(P) < t(\overline{n}+1_{H_0})$.

To discuss the second (and less trivial) case, we will again find it convenient to introduce an auxiliary element $\tilde{\delta}_{H_0,(\sigma,s,f)}$ which enables us to regard the mapping $\delta_{H_0,(\sigma,s,f)}$ as a sort of right-multiplication. Thus, define $\tilde{\delta}_{H_0,(\sigma,s,f)}\in I_{n+1}(\overline{n})$ by

$$\tilde{\delta}_{H_0,(\sigma,s,f)} = (\delta_{H_0,(\sigma,s,f)}(\mathbf{S}_0(n,\overline{n}+1_{H_0})),1,\mathbf{S}_0(n+1,\overline{n})).$$

We come now to the desired analogue of eq. (4.26), namely:

$$(\mathbf{S},g,\mathbf{S}_0)^{\sim}\cdot\tilde{\delta}_{H_0,(\sigma,s,f)} = (\delta_{H_0,(\sigma,g\cdot(s,f))}(\mathbf{S}),\tilde{g},\mathbf{S}_0) \tag{4.31}$$

for all $\mathbf{S}\in S_n(\overline{n}+1_{H_0}), g\in G(\overline{n}+1_{H_0})$, where $\tilde{g}=\sigma(g\cdot(s,f))^{-1}g\sigma(s,f)$, which is an element of $G(\overline{n})$. As in the case of eq. (4.26), this equation is also verified by looking at the picture represented by the product on the left side, noting that it does belong to $I_{n+1}(\overline{n})$, and checking that its three ingredients are indeed as given by the right side of (4.31).

Suppose now that $P=(\mathbf{Q},\rho,\mathbf{R})\in I_n(\overline{n}+1_{H_0})$. If we let $g=\rho\beta^{\mathbf{R}}_{\mathbf{T}}$, we find that $(\mathbf{Q},\rho,\mathbf{R})\cdot(\mathbf{T},1,\mathbf{S}_0) = D(\mathbf{R},\mathbf{T})(\mathbf{Q},g,\mathbf{S}_0)$, and hence

$$\begin{aligned}(D_{H_0,\sigma}(L))(P\cdot(\mathbf{T}\otimes\xi)) &= (D_{H_0,\sigma}(L))(D(\mathbf{R},\mathbf{T})(\mathbf{Q}\otimes\pi(g)\xi))\\ &= D(\mathbf{R},\mathbf{T})\sum_{(s_1,f_1)}(\delta_{H_0,(\sigma,s_1,f_1)}(\mathbf{Q})\otimes L\pi(\sigma(s_1,f_1)^{-1})\pi(g)\xi) + W^1_{n+1}(\overline{n},\chi);\end{aligned} \tag{4.32}$$

writing $(s_1,f_1)=g\cdot(s,f)$, and noting that

$$\begin{aligned}L\pi(\sigma(s_1,f_1)^{-1})\pi(g) &= L\pi(\sigma(s_1,f_1)^{-1}g)\\ &= L\pi(\sigma(s_1,f_1)^{-1}g\sigma(s,f))\pi(\sigma(s,f)^{-1})\\ &= \chi(\sigma(s_1,f_1)^{-1}g\sigma(s,f))L\pi(\sigma(s,f)^{-1}),\end{aligned}$$

(where we have used the fact that $\sigma(s_1,f_1)^{-1}g\sigma(s,f)) \in G(\overline{n})$) we see thus that the righ side of eq. (4.32) may be rewritten (after a change of variable) as

$$D(\mathbf{R},\mathbf{T})\sum_{(s,f)}(\delta_{H_0,(\sigma,g\cdot(s,f))}(\mathbf{Q})\otimes\chi(\tilde{g})L\pi(\sigma(s,f)^{-1})\xi)+W^1_{n+1}(\overline{n},\chi),$$

where $\tilde{g}=\sigma(g\cdot(s,f))^{-1}g\sigma(s,f)$, as before.

On the other hand, notice (by two applications of eq. (4.31)) that

$$\begin{aligned}(\mathbf{Q},\rho,\mathbf{R})^{\check{}}\cdot(\delta_{H_0,(\sigma,s,f)}(\mathbf{T}),1,\mathbf{S}_0) &= (\mathbf{Q},\rho,\mathbf{R})^{\check{}}\cdot(\mathbf{T},1,\mathbf{S}_0)^{\check{}}\cdot\tilde{\delta}_{H_0,(\sigma,s,f)}\\ &= ((\mathbf{Q},\rho,\mathbf{R})\cdot(\mathbf{T},1,\mathbf{S}_0))^{\check{}}\cdot\tilde{\delta}_{H_0,(\sigma,s,f)}\\ &= D(\mathbf{R},\mathbf{T})(\mathbf{Q},g,S_0)^{\check{}}\cdot\tilde{\delta}_{H_0,(\sigma,s,f)}\\ &= D(\mathbf{R},\mathbf{T})(\delta_{H_0,(\sigma,g\cdot(s,f))}(\mathbf{Q}),\tilde{g},\mathbf{S}_0).\end{aligned}$$

Hence, we see that

$$\begin{aligned}&\tilde{P}\cdot(D_{H_0,\sigma}(L))(\mathbf{T}\otimes\xi))\\ &\quad=(\mathbf{Q},\rho,\mathbf{R})^{\check{}}\cdot\sum_{(s,f)}(\delta_{H_0,(\sigma,s,f)}(\mathbf{T})\otimes L\pi(\sigma(s,f)^{-1})\xi)+W^1_{n+1}(\overline{n},\chi)\\ &\quad=D(\mathbf{R},\mathbf{T})\sum_{(s,f)}(\delta_{H_0,(\sigma,g\cdot(s,f))}(\mathbf{Q})\otimes\chi(\tilde{g})L\pi(\sigma(s,f)^{-1})\xi)+W^1_{n+1}(\overline{n},\chi),\end{aligned}$$

and the lemma is proved. □

Remark 32. (1) Fix , $\overline{n}\in N_{[n+1]}$ and $H\in\mathcal{C}$. Consider two cases now:

Case 1: $\overline{n}+1_H\in N_{[n]}$. Define $\overline{W}^2_{n+1}(\overline{n};H)$ to be the subspace of $\mathbb{C}S_{n+1}(\overline{n})/W^1_{n+1}(\overline{n}$ spanned by $\{\delta_{H,(\sigma,s,f)}(\mathbf{T})+W^1_{n+1}(\overline{n}):\mathbf{T}\in S_n(\overline{n}+1_H),1\leq s\leq\overline{n}(H)+1,f\in H\backslash N(H)$ $\sigma\in\{\sigma^H_\kappa:1\leq\kappa\leq[G:N(H)]\}\}$. By Lemma 31(2), this spanning set is a basis which ca also be characterized as $\{\mathbf{S}+W^1_{n+1}(\overline{n}):\mathbf{S}\in S_{n+1}(\overline{n}),[n+1]_{R_S}$ is not distinguished an not a singleton, and $H^S_{[n+1]}=H\}$.

Case 2: $\overline{n}+1_H\notin N_{[n]}$. In this case, define $\overline{W}^2_{n+1}(\overline{n};H)=\{0\}\subseteq\mathbb{C}S_{n+1}(\overline{n})/W^1_{n+1}(\overline{n})$. Observ that Lemma 31(2) also implies that

$$\mathbb{C}S_{n+1}(\overline{n})/W^1_{n+1}(\overline{n})=\oplus_{H\in\mathcal{C},\overline{n}+1_H\in N_{[n]}}\overline{W}^2_{n+1}(\overline{n};H).$$

(2) We will need the following slightly stronger version of Remark 30(2): Let $\mathcal{G}_0$ be subgroup of a finite group $\mathcal{G}$ and $\mathcal{G}=\coprod_{k=1}^n g_k\mathcal{G}_0$ with $g_1=1$. For $\chi\in\hat{\mathcal{G}}_0$ and $\tilde{\chi}=\mathrm{Ind}_{\mathcal{G}_0\uparrow\mathcal{G}}(\chi$ we may regard $V_{\tilde{\chi}}$ as the space $V_\chi\otimes\mathbb{C}(\mathcal{G}/\mathcal{G}_0)$ with $\mathcal{G}$-action defined by $g(v\otimes g_i\mathcal{G}_0)$ $\chi(h(g,i))(v)\otimes g_{g(i)}\mathcal{G}_0$ where $gg_i=g_{g(i)}h(g,i)$ with $g_{g(i)}\in\{g_1,\ldots,g_k\}$ and $h(g,i)\in\mathcal{G}$ Furthermore, for $\pi\in\hat{\mathcal{G}}$, there is a natural bijection between $\mathcal{G}_0$-linear maps $L:V_\pi\longrightarrow V$ and $\mathcal{G}$-linear maps $\tilde{L}:V_\pi\longrightarrow V_{\tilde{\chi}}$ given by $\tilde{L}(\xi)=\sum_{k=1}^n L\pi(g_k^{-1})(\xi)\otimes g_k\mathcal{G}_0$.

(3) Fix $H\in\mathcal{C}$ such that $\overline{n}+1_H\in N_{[n]}$; also fix $\chi\in\widehat{G(\overline{n})}$, and let $\tilde{\chi}$ be the result inducing χ up to $G(\overline{n}+1_H)$. For appropriate $\pi\in\widehat{G(\overline{n}+1_H)}$, pick non-zero $G(\overline{n}+1_H$ linear maps $\tilde{L}_{\pi,j}:V_\pi\longrightarrow V_{\tilde{\chi}}$ such that their ranges afford a direct sum decomposition $V_{\tilde{\chi}}$. Let $L_{\pi,j}$ be related to $\tilde{L}_{\pi,j}$ as in (2) above. Then, by an argument exactly analogous the corresponding one used in Remark 30(3), it may be verified that:

$$\oplus_{\sigma,\pi,j}\text{ range }D_{H,\sigma}(L_{\pi,j})=\overline{W}^2_{n+1}(\overline{n};H)\otimes V_\chi. \qquad (4.3$$

(4) We conclude from (1) and (3) that

$$V_{n+1}(\overline{n},\chi)/W^1_{n+1}(\overline{n},\chi) = \oplus_{H\in C,\overline{n}+1_H\in N_{[n]}} \oplus_{\sigma,\pi,j} \text{ range } D_{H,\sigma}(L_{\pi,j})$$

which is a decomposition of the left hand side as a direct sum of irreducible $A^G_n(d)$-modules.

All the pieces are now in place for the required description of the Bratteli diagram for the inclusion $A^G_n(d) \subset A^G_{n+1}(d)$.

Theorem 33. *Fix $n \in \mathbb{N}$ and let d be any positive number satisfying the hypothesis of Theorem 21. Then,*

(a) *The set $\widehat{A^G_n(d)}$ of irreducible representations of $A^G_n(d)$ can be parametrised by the set $\{(n,\overline{n},\pi) : \overline{n} \in N_{[n]}, \pi \in \widehat{G(\overline{n})}\}$, in such a way that the associated module $V_n(\overline{n},\pi)$ has dimension equal to $|S_n(\overline{n})|d_\pi$.*

(b) *When viewed as an $A^G_n(d)$-module, the multiplicity with which the module $V_{n+1}(\overline{n},\chi)$ contains $V_n(\overline{m},\pi)$ is given by:*

  (i) $\langle\chi|_{G(\overline{m})},\pi\rangle[G:N(H)]$, *if $\overline{n}=\overline{m}+1_H$ for some $H \in C$;*

  (ii) $\delta_{\pi,\chi}\sum_{H\in C}[G:N(H)] + \sum_{H\in C,\overline{n}(H)>0}[G:N(H)]\langle\tilde{\chi},\pi\rangle$, *if $\overline{n}=\overline{m}$, where $\tilde{\chi} = \text{Ind}_{G(\overline{n}-1_H)\uparrow G(\overline{n})}\text{Res}_{G(\overline{n})\downarrow G(\overline{n}-1_H)}\chi$;*

  (iii) $\langle\pi|_{G(\overline{n})},\chi\rangle[G:N(H)]$, *if $\overline{n}=\overline{m}-1_H$, for some $H \in C$; and*

  (iv) 0, *otherwise.*

Proof. (a) is an immediate consequence of Theorem 21, while (b) follows from Remark 27(4), Remark 30(4), Remark 32(4), and the simple fact that if W is a sub-module of a semi-simple module V, then $V \cong W \oplus (V/W)$. □

5. Concluding remarks

We wish to make a few remarks in two directions: first, we wish to discuss the special case of the algebra $A^G_1(d)$ and the question of whether, as a filtered algebra (with filtering given by the number of through classes), this determines the group G; and second, we wish to discuss the trivial specialization $|G| = 1$, which has already appeared in the literature.

The filtered algebra $A^G_1(d)$

The algebra $A^G_1(d)$ admits the filtration given by the ideals $I^G_k, 0 \leq k \leq |G|$. Further, it should be clear from the definitions that if $k \neq 0$, then $N_{1;k} \neq \emptyset$ only when $k|n$, in which case, $\overline{n} \in N_{1;k}$ if and only if there exists $H_0 \in C$ such that $[G : H_0] = k$ and $\overline{n}(H) = \delta_{H,H_0}$. In particular, if we assume, further, that G is abelian, then, for any fixed divisor k of $|G|$, we see from Theorem 21 that

$$I^G_k/I^G_{k-1} \cong \oplus_H(\mathbb{C}^k \otimes M_k(\mathbb{C})),$$

where the direct sum is over all subgroups H of G of index k. In particular, the knowledge of the filtered algebra $A^G_1(d)$ amounts, in case G is abelian, to the knowledge of the number $s_G(l)$ of subgroups of G of any given order l. It is a curious fact that this knowledge of an abelian group G – i.e., of the function $s_G : \mathbb{N} \to \mathbb{N} \cup \{0\}$ defined by the previous sentence – completely determines the abelian group G (up to isomorphism).

It is natural to ask whether the filtered algebra $A_1^G(d)$ determines the isomorphism cla of the (general, possibly non-abelian) group G. What we can see is that the filter algebra $A_1^G(d)$ determines whether or not G is simple, and more generally, it can dete the set of orders of all quotients of G.

The case $|G| = 1$

When $G = \{1\}$ is the trivial group, then what we have called $A_n^G(d)$ is exactly the same what we called $A_n(d)$; this algebra was originally discussed in [J], where it was shov that at least when $d = k$, the algebra $A_n(d)$ can be identified with the commutant of t natural representation (given by the diagonal action) of Σ_k on $\otimes^n V$, where V is a dimensional vector space, provided that $k \geq 2n$.

Algebras related to certain subalgebras of the general $A_n(d)$ have occurred in numero contexts – for instance, see [B, J, W]; (see also Remark 2).

Recall that the Temperley–Lieb algebra is the subalgebra of $A_n(d)$ generated by tho equivalence relations which have the property that all equivalence classes are tw element sets, and whose diagrams are planar. Another subalgebra, call it $B_n(d)$, obtained by dropping this planarity requirement. The structure of the inclusion $B_n \subset$ may be analysed using techniques similar to the ones discussed here, and involve t representation theory of the various symmetric groups and the 'induction-restrictio relations between several naturally arising subgroups thereof.

Acknowledgement

We would like to record our gratitude to Dr. Shreedhar Inamdar for useful conversatio particularly in connection with the identification of the statement and proof of the use Lemma 19.

References

[B] Bhattacharyya B, *Krishnan-Sunder subfactors and a new countable family of subfact related to trees*, Ph.D. thesis, UC Berkeley

[J] Jones V F R, *The Potts model and the symmetric group*, Subfactors (Kyuzeso, 1993) (N Jersey: World Sci. Pub.) (1994) pp. 259–267

[J1] Jones V F R, *Planar Algebras*, preprint (1997)

[M] Martin Paul, Temperley-Lieb algebras for non-planar statistical mechanics – the parti algebra construction, *J. Knot Theory Ramifications* **3** (1994) 51–82

[M1] Martin Paul, The structure of the partition algebras, *J. Algebra* **183** (1996) 319–358

[W] Wenzl H, On the structure of Brauer's centraliser algebras, *Ann. Math.* **128** 173–193 (19

Proc. Indian Acad. Sci. (Math. Sci.), Vol. 110, No. 3, August 2000, pp. 293–304.
© Printed in India

On the generalized Hankel–Clifford transformation of arbitrary order

S P MALGONDE and S R BANDEWAR

Department of Mathematics, College of Engineering, Kopargaon 423 603, India
Email: sescolk@giaspnol.vsnl.in

MS received 24 November 1999; revised 6 March 2000

Abstract. Two generalized Hankel–Clifford integral transformations verifying a mixed Parseval relation are investigated on certain spaces of generalized functions for any real value of their orders $(\alpha - \beta)$.

Keywords. Generalized Hankel–Clifford transformation; Parseval relation; generalized functions (distributions).

1. Introduction

The conventional Hankel transformation defined by

$$h_\mu\{f(x)\}(y) = \int_0^\infty \sqrt{xy}\, J_\mu(xy) f(x)\mathrm{d}x \quad (0 < y < \infty) \tag{1}$$

was extended by [6] to certain generalized function of slow growth through a generalization of Parseval's equation. Later on [1] extended (1) to a class of generalized functions by the kernel method, which is a more natural extension of (1). In [3] the Hankel–Clifford transformations of order $\mu \geq 0$, defined by

$$(h_{1,\mu}f)(y) = y^\mu \int_0^\infty (xy)^{-\mu/2} J_\mu(2\sqrt{xy}) f(x)\mathrm{d}x \tag{2}$$

and

$$(h_{2,\mu}f)(\,) = \int_0^\infty x^\mu (xy)^{-\mu/2} J_\mu(2\sqrt{xy})(x)\mathrm{d}x \tag{3}$$

has been extended to a certain space of generalized functions. Recently the simple generalization of (2) and (3) called the generalized Hankel–Clifford transformation of order $\alpha - \beta \geq -1/2$ are defined by

$$F_1(y) = (h_{1,\alpha,\beta}f)(y) = y^{-\alpha-\beta} \int_0^\infty C_{\alpha,\beta}(xy) f(x)\mathrm{d}x, \tag{4}$$

$$F_2(y) = (h_{2,\alpha,\beta}f)(y) = \int_0^\infty x^{-\alpha-\beta} C_{\alpha,\beta}(xy) f(x)\mathrm{d}x, \tag{5}$$

where $C_{\alpha,\beta}(xy) = (xy)^{(\alpha+\beta)/2} J_{\alpha-\beta}(2\sqrt{xy})$ is a generalized Bessel–Clifford function of first kind of order $(\alpha - \beta)$, satisfying the differential equation $xy'' - (\alpha + \beta - 1)y' + (\alpha\beta x^{-1} + 1)y = 0$, which can be inverted by formulae

$$f(y) = (h^{-1}_{1,\alpha,\beta}F_1)(y) = y^{-\alpha-\beta}\int_0^\infty C_{\alpha,\beta}(xy)F_1(x)\mathrm{d}x,$$

$$f(y) = (h^{-1}_{2,\alpha,\beta}F_2)(y) = \int_0^\infty x^{-\alpha-\beta}C_{\alpha,\beta}(xy)F_2(x)\mathrm{d}x,$$

respectively. In symbols $h^{-1}_{1,\alpha,\beta} = h_{1,\alpha,\beta}$ and $h^{-1}_{2,\alpha,\beta} = h_{2,\alpha,\beta}$, if $\alpha - \beta \geq -1/2$. T distributional aspects of the transformations (6) and (7) have been discussed by t kernel method in [2].

In this paper the two transformations (4) and (5) are simultaneously investigated certain spaces of distributions, by the procedure of the adjoint operator for which exploit a Parseval equation involving (4) and (5) so that we define the first distributio transformation as the adjoint operator of the second one, and conversely for $(\alpha-\beta) \geq -1$ as well as for negative real values of order $(\alpha - \beta)$ similar to [4] and [5].

2. Preliminary results

We shall need the operational formulae given by Malgonde [2]

$$D^r[x^{-\alpha}C_{\alpha,\beta}(x)] = (-1)^r x^{-\alpha}C_{\alpha,\beta-r}(x),$$
$$D^r[x^{-(\beta-r)}C_{\alpha,\beta-r}(x)] = x^{-\beta}C_{\alpha,\beta}(x),$$

and the asymptotic behaviours

$$C_{\alpha,\beta}(x) = O(|x|^\alpha) \text{ as } x \longrightarrow 0^+$$

and

$$C_{\alpha,\beta}(x) = O(x^{(\alpha+\beta)/2-1/4}) \text{ as } x \longrightarrow \infty.$$

Under certain assumptions, the Parseval equation for (4) given by

$$\int_0^\infty x^{\alpha+\beta}f(x)g(x)\mathrm{d}x = \int_0^\infty y^{\alpha+\beta}F_1(y)G_1(y)\mathrm{d}y$$

holds.

The corresponding Parseval equation for (5) takes the form

$$\int_0^\infty x^{-(\alpha+\beta)}f(x)g(x)\mathrm{d}x = \int_0^\infty y^{-(\alpha+\beta)}F_2(y)G_2(y)\mathrm{d}y.$$

Note that (4) and (5) are two variants of the Hankel transformation in the form conside by Tricomi. For $\alpha = 0$, $\beta = -\mu$ these reduce to Hankel–Clifford transformations gi by (2) and (3). In the sequel I denotes the interval $0 < x < \infty$ and $L(I)$ represents space of all functions $f(x)$ that are integrable Lebesgue on I. By invoking Fubi theorem we can establish the following theorem.

Theorem 1. *Let $\alpha - \beta \geq -1/2$, suppose $x^\alpha f(x)$ and $y^\beta G_2(Y)$ belongs to $L(I)$ $F_1(y) = (h_{1,\alpha,\beta}f)(y)$ and $g(x) = (h^{-1}_{2,\alpha,\beta}G_2(y))(x)$ then*

$$\int_0^\infty f(x)g(x)\mathrm{d}x = \int_0^\infty F_1(y)G_2(y)\mathrm{d}y.$$

Note that (14) does not involve any weight function, in contrast with expressions (12) and (13). Moreover, (14) relates both the Hankel–Clifford transforms (4) and (5). For this reason, (14) will be named as the mixed Parseval equation concerning generalized Hankel–Clifford transforms (4) and (5), and will play an important role throughout this paper.

Along this paper we follow the notation and terminology of Zemanian [6]. Thus $D(I)$, $E(I)$, $D'(I)$ and $E'(I)$ denote well known testing function spaces and their duals.

3. The testing function spaces $\mathbb{H}_\beta(I)$, $\mathbb{S}_\alpha(I)$ and their duals

Let β by any real number. $\mathbb{H}_\beta(I)$ is the space of all infinitely differentiable functions defined on I such that

$$\gamma^{\beta}_{m,k}(\phi) = \sup_{x\in I} |x^m D^k x^\beta \phi(x)| \tag{15}$$

exists for each pair of non-negative integers m and k. We equip with this space topology generated by the countable multinorm (15). Thus $\mathbb{H}_\beta(I)$ is a Frechet space.

If $\phi(x)$ admits the expansion

$$\phi(x) = x^{-\beta}[a_0 + a_1x + a_2x^2 + \cdots + a_px^p + O(x^p)] \tag{16}$$

near the origin, a result analogous to Zemanian [6] can be deduced.

Lemma 2.1. *$\phi(x)$ is a member of $\mathbb{H}_\beta(I)$ if and only if $\phi(x)$ is an infinitely differentiable function, $\phi(x)$ has the form* (16) *in some vicinity of the origin and $D^k\phi(x)$ is of rapid descent when $x \to \infty$ for each $k = 0, 1, 2, \ldots$.*

$\mathbb{H}'_\beta(I)$ is the dual space of $\mathbb{H}_\beta(I)$. We consider only in $\mathbb{H}'_\beta(I)$ the weak topology ([6], p. 21). $\mathbb{H}'_\beta(I)$ is too complete. We point out the properties:

(i) The inclusions $D(I) \subset \mathbb{H}_\beta(I) \subset E(I)$ hold. $E'(I)$ is a subspace of $\mathbb{H}'_\beta(I)$.

(ii) The mapping $x^{-1}\phi(x) \to \phi(x)$ is an isomorphism from $\mathbb{H}_{\beta-1}(I)$ onto $\mathbb{H}_\beta(I)$. Therefore the operator $f(x) \to x^{-1}f(x)$ defined by

$$\langle x^{-1}f(x),\ \phi(x)\rangle = \langle f(x), x^{-1}\phi(x)\rangle,$$

$f \in \mathbb{H}'_\beta(I)$, $\phi(x) \in \mathbb{H}'_{\beta-1}(I)$ is an isomorphism from $\mathbb{H}'_\beta(I)$ on to $\mathbb{H}'_{\beta-1}(I)$.

(iii) $\mathbb{H}_{\beta-r}(I)$ is a subspace of $\mathbb{H}_\beta(I)$, for any position integer r.

On the other hand, $\mathbb{S}_\alpha(I)$ consists of all infinitely differentiable functions $\psi(x)$ defined on I so that

$$\rho^{\alpha}_{m,k}(\psi) = \sup_{x\in I} |x^m D^k x^{-\alpha}\psi(x)| \tag{17}$$

exists for each pair of non-negative integers m and k. With the topology generated by (17), $S_\alpha(I)$ is also a Frechet space. Assume that $\psi(x)$ takes the form

$$\psi(x) = x^\alpha[b_0 + b_1x + b_2x^2 + \cdots + b_px^p + O(x^p)] \tag{18}$$

as $x \to 0^+$. We can then prove the following lemma.

Lemma 2.2. *$\psi(x)$ belongs to $\mathbb{S}_\alpha(I)$ if and only if $\psi(x)$ is infinitely differentiable, $\psi(x)$ is of the form* (18) *as $x \to 0^+$ and $D^k\psi(x)$ is of rapid descent as $x \to \infty$, for each $k = 1, 2, 3, \ldots$.*

(iv) The inclusions $D(I) \subset \mathbb{S}_\alpha(I) \subset E(I)$ hold. $E'(I)$ is a subspace of $S'_\alpha(I)$.

(v) The mapping $\psi(x) \to x\psi(x)$ is an isomorphism from $\mathbb{S}_\alpha(I)$ onto $\mathbb{S}_{\alpha+1}(I)$. Therefo $f(x) \to xf(x)$ defined by $\langle xf(x), \Psi(x)\rangle = \langle f(x), x\psi(x)\rangle$, $f \in \mathbb{S}'_{\alpha+1}(I)$, $\psi(x) \in \mathbb{S}_\alpha($ is an isomorphism from $\mathbb{S}'_{\alpha+1}(I)$ onto $\mathbb{S}'_\alpha(I)$.

(vi) $\mathbb{S}_{\alpha+r}(I)$ is a subspace of $\mathbb{S}_\alpha(I)$, for any positive integer r.

We next discuss the following differential operators:

$$R_\beta = x^{-\beta+1}Dx^\beta, \qquad (1$$

$$Q_\alpha = x^{-\alpha}Dx^\alpha, \qquad (2$$

$$R^*_\beta = -x^\beta Dx^{-\beta+1}, \qquad (2$$

$$Q^*_\alpha = -x^\alpha Dx^{-\alpha}. \qquad (2$$

When these operators act on these spaces $\mathbb{H}_\beta(I)$ and $\mathbb{S}_\alpha(I)$, one verifies the followi lemma.

Lemma 2.3. (a) *The differential operator R_β is an isomorphism from $\mathbb{H}_\beta(I)$ into $\mathbb{H}_{\beta-1}($ its inverse being*

$$R_\beta^{-1} = x^{-\beta}\int_\infty^x t^{\beta-1}\phi(t)\mathrm{d}t, \quad \phi \in \mathbb{H}_{\beta-1}(I). \qquad (2$$

(b) *The operator Q_α is a continuous linear mapping from $\mathbb{H}_{\beta-1}(I)$ into $\mathbb{H}_\beta(I)$.* (c) *T differential operator R^*_β is a continuous linear mapping from $\mathbb{S}_\alpha(I)$ into $\mathbb{S}_\alpha(I)$.* (d) *T differential operator Q^*_α is an automorphism from $\mathbb{S}_\alpha(I)$ onto $\mathbb{S}_\alpha(I)$, its inverse bein*

$$(Q^*_\alpha)^{-1} = -x^\alpha\int_\infty^x t^{-\alpha}\phi(t)\mathrm{d}t. \qquad ($$

Proof. Let $\phi \in \mathbb{H}_\beta(I)$. It can be seen that

$$\gamma^{\beta-1}_{m,k}[R_\beta\phi] = \gamma^\beta_{m,k+1}[\phi]. \qquad ($$

On the other hand, if $\phi \in \mathbb{H}_{\beta-1}(I)$ we have

$$\gamma^\beta_{m,k}[R_\beta^{-1}\phi] = \gamma^{\beta-1}_{m,k-1}[\phi], \qquad ($$

$k = 1, 2, 3, \ldots$ and for the case $k = 0$,

$$\gamma^\beta_{m,o}[R_\beta^{-1}\phi] \leq \frac{\pi}{2}[\gamma^{\beta-1}_{m,o}(\phi) + \gamma^{\beta-1}_{m+2,o}[\phi]]. \qquad ($$

To verify (b), note that

$$\gamma^\beta_{m,k}[Q_\alpha(\phi)] \leq |\alpha - \beta + k + 1|\gamma^{\beta-1}_{m,k}[\phi] + \gamma^{\beta-1}_{m+1,k+1}[\phi]$$

holds for every $\phi \in \mathbb{H}_{\beta-1}(I)$.

A similar reasoning permits one to prove (c) and (d). But, if the same operators considered acting on the spaces of generalized functions $\mathbb{H}'_\beta(I)$ and $\mathbb{S}'_\alpha(I)$, the follow assertions can be derived as an immediate consequences of Lemma 2.3.

Lemma 2.4. (a$'$) *The generalized operator* R^*_β, *defined on* $\mathbb{H}'_{\beta-1}(I)$ *as the adjoint of* R_β, *that is*, $\langle R^*_\beta f, \phi\rangle = \langle f, R_\beta\phi\rangle$, $f \in \mathbb{H}'_{\beta-1}(I)$, $\phi \in \mathbb{H}_\beta(I)$, *is an isomorphism from* $\mathbb{H}'_\beta(I)$ *onto* $H'_{\beta-1}(I)$. (b$'$) *The generalized operator* Q^*_α *defined in* $\mathbb{H}'_\beta(I)$, *as usual by* $\langle Q^*_\alpha f, \phi\rangle = \langle f, Q_\alpha\phi\rangle$, $f \in \mathbb{H}'_\beta(I)$, $\phi \in \mathbb{H}_{\beta-1}(I)$, *is a continuous linear mapping of* $\mathbb{H}'_\beta(I)$ *into* $H'_{\beta-1}(I)$. (c$'$) *The generalized operator* R_β, *defined in* $\mathbb{S}'_\alpha(I)$ *as the adjoint of* R^*_β *in* $\mathbb{S}_\alpha(I)$, *namely*, $\langle R_\beta f, \phi\rangle = \langle f, R^*_\beta\phi\rangle$, $f \in \mathbb{S}'_\alpha(I)$, $\phi \in \mathbb{S}_\alpha(I)$, *is a continuous linear operator on* $\mathbb{S}'_\alpha(I)$ *into itself.* (d$'$) *The generalized operator* Q_α *defined in* $\mathbb{S}'_\alpha(I)$ *through* $\langle Q_\alpha f, \phi\rangle = \langle f, Q^*_\alpha\phi\rangle$, $f \in \mathbb{S}'_\alpha(I)$, $\phi \in \mathbb{S}_\alpha(I)$, *is an automorphism on* $\mathbb{S}'_\alpha(I)$.

Now assume that $\alpha - \beta \geq -1/2$, then $\mathbb{H}_\beta(I)$ can be identified with a subspace of $\mathbb{S}'_\alpha(I)$. Indeed, every member of $f \in \mathbb{H}_\beta(I)$ gives rise to a regular generalized function $f \in \mathbb{S}'_\alpha(I)$ by

$$\langle f, \phi\rangle = \int_0^\infty f(x)\phi(x)\mathrm{d}x, \quad \phi \in \mathbb{S}\alpha(I),$$

since the linear operator f satisfies

$$|\langle f, \theta\rangle| \leq \rho^\alpha_{0,0}(\phi)\int_0^\infty |x^{-\alpha}f(x)|\mathrm{d}x$$

for all $\phi \in \mathbb{S}_\alpha(I)$ and $f(x)$ is a function of rapid descent, that is, f is also continuous. Observe that the last integral exists is view of Lemma 2.1, because $f \in \mathbb{H}_\beta(I)$. Moreover, two members of $\mathbb{H}_\beta(I)$ which generates the same regular distribution in $\mathbb{S}'_\alpha(I)$ must be identical. These consideration justify the inclusion $\mathbb{H}_\beta(I) \subset \mathbb{S}'_\alpha(I)$. We can argue in a similar way to verify that $\mathbb{S}_\alpha(I) \subset \mathbb{H}'_\beta(I)$.

Remark 1. The generalized Kepinski operator, as given in [2] $K_{\alpha,\beta} = xD^2 - (\alpha + \beta - 1)$ $D + \alpha\beta x^{-1} = R^*_\beta Q^*_\alpha$ is a continuous linear mapping from the spaces $\mathbb{S}_\alpha(I)$ and $\mathbb{H}'_\beta(I)$ into themselves. Analogously, $K_{\alpha,\beta} = xD^2 + (\alpha + \beta + 1)D + \alpha\beta x^{-1} = Q_\alpha R_\beta$ is a continuous linear mapping from the spaces $\mathbb{H}_\beta(I)$ and $\mathbb{S}'_\alpha(I)$ into themselves.

4. The classical generalized Hankel–Clifford transformations on the space $\mathbb{H}_\beta(I)$ and $\mathbb{S}_\alpha(I)$

Under the restriction $(\alpha - \beta) \geq -1/2$ every member of $\mathbb{H}_\beta(I)$ fulfils the requirements of inversion theorem of (4) and consequently, the generalized Hankel–Clifford transformation $h_{1,\alpha,\beta}$ given by

$$(h_{1,\alpha,\beta}\phi)(y) = \Phi(y) = y^{-(\alpha+\beta)}\int_0^\infty C_{\alpha,\beta}(xy)\phi(x)\mathrm{d}x \tag{28}$$

exists for all $\phi \in \mathbb{H}_\beta(I)$. We next establish the main result.

Theorem 2. *Let* $(\alpha - \beta) \geq -1/2$. *The first generalized Hankel–Clifford transformation* $h_{1,\alpha,\beta}$ *is an automorphism on* $\mathbb{H}_\beta(I)$.

Proof. Let $\phi(x)$ be any member of $\mathbb{H}_\beta(I)$. Expression (28) defines obviously a linear operator on $\mathbb{H}_\beta(I)$. To prove its continuity, note that the smoothness of the integrand and the use of (8) and (9) yield

$$y^m D_y^k y^\beta \Phi(y) = y^m D_y^k y^\beta \left[y^{-(\alpha+\beta)} \int_0^\infty C_{\alpha,\beta}(xy)\phi(x)\mathrm{d}x \right]$$
$$= (-1)^k y^m \int_0^\infty \frac{\mathrm{d}^N}{\mathrm{d}x^N}(x^{-\beta+k+N} C_{\alpha+k+N,\beta}(xy)) x^\beta \phi(x)\mathrm{d}x$$

where N denotes an arbitrary positive integer. If we set $N = 2m$ and integrate by parts times in the last integral, one can deduce

$$y^m D_y^k y^\beta \Phi(y) = \int_0^\infty (xy)^m C_{\alpha+k+2m,\beta}(xy) x^{-\beta+k+m} D_x^{2m} x^{(-\alpha-\beta)}\phi(x)\mathrm{d}x \qquad (2$$

since $|z^m C_{\alpha+k+2m,\beta}(z)| \leq A_{m,k}$ on $0 < z < \infty$, $A_{m,k}$ being a positive constant. If represents a positive integer no less than $-\beta + k + m$, it requires from (29) that

$$|y^m D_y^k y^\beta \Phi(y)| \leq A_{m,k} \sum_{r=0}^{n+2} \binom{n+2}{r} \nu_{r,2m}^\beta(\phi) \int_0^\infty \frac{\mathrm{d}x}{(1+x)^2},$$

that is

$$\nu_{m,k}^\beta(\phi) \leq A_{m,k} \sum_{r=0}^{n+2} \binom{n+2}{r} \nu_{r,2m}^\beta(\phi).$$

Expression (29) shows that $\Phi(y)$ is an infinitely differentiable function, whereas (implies that $h_{1,\alpha,\beta}$ is a continuous operator on $\mathbb{H}_\beta(I)$. Finally, by inversion theorem [2] have $h_{1,\alpha,\beta}^2\phi = \phi$ for all $\phi \in \mathbb{H}_\beta(I)$. This completes the proof of Theorem 2.

The second Hankel–Clifford transformation $h_{2,\alpha,\beta}$ defined by means of

$$(h_{2,\alpha,\beta}\psi)(y) = \Psi(y) = \int_0^\infty x^{-(\alpha+\beta)} C_{\alpha,\beta}(xy)\psi(x)\mathrm{d}x$$

exists for every $\psi \in \mathbb{S}_\alpha(I)$, by inversion theorem [2]. Through an argument similar to one used in the proof of Theorem 2, we can assert the next theorem.

Theorem 3. *The second generalized Hankel–Clifford transformation $h_{2,\alpha,\beta}$ of or $(\alpha - \beta) \geq -1/2$ is an automorphism on $\mathbb{S}_\alpha(I)$.*

Now some interesting operational rules for the transformation $h_{1,\alpha,\beta}$ are obtained.

Lemma 4.1. *Let $\alpha - \beta \geq -1/2$. For all $\phi(x) \in \mathbb{H}_\beta(I)$, we have*

$$R_\beta h_{1,\alpha,\beta}(\phi) = h_{1,\alpha,\beta-1}(-x\phi),$$
$$h_{1,\alpha,\beta-1}(R_\beta\phi) = -y h_{1,\alpha,\beta}(\phi),$$
$$h_{1,\alpha,\beta}(Q_\alpha R_\beta\phi) = -y h_{1,\alpha,\beta}(\phi),$$
$$Q_\alpha R_\beta h_{1,\alpha,\beta}(\phi) = h_{1,\alpha,\beta}(-x\phi),$$

and for all $\phi(x) \in \mathbb{H}_{\beta-1}(I)$

$$h_{1,\alpha,\beta}(Q_\alpha\phi) = h_{1,\alpha,\beta-1}(\phi),$$
$$Q_\alpha h_{1,\alpha,\beta-1}(\phi) = h_{1,\alpha,\beta}(\phi).$$

Proof. Let $\phi \in \mathbb{H}_\beta(I)$. We may differentiate under the integral sign to obtain, in accordance with (8),

$$R_\beta h_{1,\alpha,\beta}\phi = y^{-\alpha-\beta+1} \int_0^\infty (-x)C_{\alpha,\beta-1}(xy)\phi(x)\mathrm{d}x = h_{1,\alpha,\beta-1}(-x\phi).$$

This proves (32). To see (33), the integration by parts and use of (9) yield

$$\begin{aligned} h_{1,\alpha,\beta-1}(R_\beta\phi) &= y^{-(\alpha+\beta-1)} \int_0^\infty x^{-\beta+1} C_{\alpha,\beta-1}(xy) D_x x^\beta \phi(x)\mathrm{d}x \\ &= y^{-(\alpha+\beta-1)} \left\{ (xC_{\alpha,\beta-1}(xy)\phi)_{x\to 0}^{x\to\infty} + - \int_0^\infty x_0^{-\beta} C_{\alpha,\beta-1}(xy) D_x^\beta x\phi(x)\mathrm{d}x \right\}. \end{aligned}$$

The limit terms tend to zero since $\phi(x)$ is of rapid descent as $x \to \infty$ and, as $x \to 0^+$, $xC_{\alpha,\beta-1}(xy) = \infty(|x|^{\alpha+1})$ and $\phi(x) = 0(|x|^{-\beta})$ when $(\alpha - \beta) \geq -1/2$, this verifies (33). Similar manipulation allow us to obtain the remaining formulas.

Next we summarize operational calculus generated by the transform $h_{2,\alpha,\beta}$.

Lemma 4.2. *Let* $\alpha - \beta \geq -1/2$. *For all* $\psi(x) \in \mathbb{S}_\alpha(I)$, *we have*

$$Q_\alpha^* h_{2,\alpha,\beta}(\psi) = h_{2,\alpha,\beta-1}(\psi), \tag{38}$$

$$h_{2,\alpha,\beta-1}(Q_\alpha^*\psi) = h_{2,\alpha,\beta}(\psi), \tag{39}$$

$$h_{2,\alpha,\beta}(R_\beta^* Q_\alpha^*\psi) = -yh_{2,\alpha,\beta-1}(\psi), \tag{40}$$

$$R_\beta^* Q_\alpha^* h_{2,\alpha,\beta}(\psi) = h_{2,\alpha,\beta}(-x\psi). \tag{41}$$

For $\psi(x) \in \mathbb{S}_{\beta-1}(I)$,

$$h_{2,\alpha,\beta}(R_\beta^*\psi) = -yh_{2,\alpha,\beta-1}(\psi), \tag{42}$$

$$R_\alpha^* h_{2,\alpha,\beta-1}(\psi) = h_{2,\alpha,\beta-1}(-x\psi). \tag{43}$$

5. The distributional generalized Hankel–Clifford transformation $h'_{1,\alpha,\beta}$

Let $(\alpha - \beta) \geq -1/2$. We propose defining the distributional generalized Hankel–Clifford transformation $h'_{1,\alpha,\beta}$ on $\mathbb{S}'_\alpha(I)$ as the adjoint operator of $h_{2,\alpha,\beta}$ on $\mathbb{S}_\alpha(I)$, that is

$$\langle h'_{1,\alpha,\beta} f, \Phi\rangle = \langle f, h_{2,\alpha,\beta}\Phi\rangle \tag{44}$$

for all $f \in \mathbb{S}'_\alpha(I)$ and $\Phi \in \mathbb{S}_\alpha(I)$.

Note that by setting $\Phi = h_{2,\alpha,\beta}\phi$, $\phi \in \mathbb{S}_\alpha(I)$, (44) takes the form

$$\langle h'_{1,\alpha,\beta} f, h_{2,\alpha,\beta}\phi\rangle = \langle f, \phi\rangle, \quad f \in \mathbb{S}'_\alpha(I), \quad \phi \in \mathbb{S}_\alpha(I). \tag{45}$$

Hence (45) can be understood as a generalization of the mixed Parseval equation (14), as it happens in the extention of the Hankel transform to certain space of generalized functions ([6], p. 142). Note that definition (44) has a sense. Indeed, from Theorem 2 the following is inferred.

Theorem 3. *Let* $(\alpha - \beta) \geq -1/2$. *The distributional generalized Hankel–Clifford transformation* $h'_{1,\alpha,\beta}$, *as defined by* (44), *is an automorphism on* $\mathbb{S}'_\alpha(I)$.

Recall that $\mathbb{H}_\beta(I) \subset \mathbb{S}'_\alpha(I)$. If $f \in \mathbb{H}_\beta(I)$, then $h'_{1,\alpha,\beta}f$ exists and f gives rise to a regular member on $\mathbb{S}'_\alpha(I)$. We may write by (44)

$$\langle h'_{1,\alpha,\beta}f, \Phi\rangle = \langle f, h_{2,\alpha,\beta}\Phi\rangle = \int_0^\infty f(x)(h_{2,\alpha,\beta}\Phi)(x)\mathrm{d}x$$

for all $\Phi \in \mathbb{S}_\alpha(I)$. By applying (14) we can convert last integral into

$$\int_0^\infty (h_{1,\alpha,\beta}f)(y)\Phi(y)\mathrm{d}y = \langle h_{1,\alpha,\beta}f, \Phi\rangle.$$

Therefore, whatever $f \in \mathbb{H}_\beta(I)$

$$h'_{1,\alpha,\beta}f = h_{1,\alpha,\beta}f, \tag{46}$$

that is to say, the classical transformation $h_{1,\alpha,\beta}$ is a special case of the distributional transformation $h'_{1,\alpha,\beta}$ given by (44). From Lemma 4.2 we deduce immediately, in line with new definition (44), the following operational formulas.

Lemma 5.1. *Let* $(\alpha-\beta) \geq -1/2$. *For all* $f(x) \in \mathbb{S}'(I)$, *we have*

$$\begin{aligned} R_\beta h'_{1,\alpha,\beta}(f) &= h'_{1,\alpha,\beta-1}(-xf),\\ h'_{1,\alpha,\beta-1}(R_\beta f) &= -yh'_{1,\alpha,\beta}(f),\\ h'_{1,\alpha,\beta}(Q_\alpha R_\beta f) &= -yh'_{1,\alpha,\beta}(f),\\ Q_\alpha R_\beta h'_{1,\alpha,\beta}(f) &= h'_{1,\alpha,\beta}(-xf), \end{aligned}$$

and for all $f(x) \in \mathbb{S}_{\alpha+1}(I)$

$$\begin{aligned} h'_{1,\alpha,\beta}(Q_\alpha f) &= h'_{1,\alpha,\beta-1}(f),\\ Q_\alpha h'_{1,\alpha,\beta-1}(f) &= h'_{1,\alpha,\beta}(f). \end{aligned}$$

Since $\mathbb{H}_\beta(I) \subset \mathbb{S}'_\alpha(I)$ *and* $\mathbb{H}_{\beta-1}(I) \subset \mathbb{S}'_\alpha(I)$, *these results generalize those in Lemma* 4.1.

Note that the classical operational formulae coincide with their respective distributional expressions.

$\mathbb{H}_{\alpha,\beta,a}$ is the space of all infinitely differentiable functions $\phi(x)$ defined on I for which

$$\gamma_k^{\alpha,\beta,a}(\phi) = \sup_{x\in I}|\mathrm{e}^{-ax}x^{-\alpha}\Delta^k_{\alpha,\beta}\phi(x)| < \infty, \quad k = 0,1,2,\ldots, \tag{47}$$

where $a > 0$, $(\alpha-\beta) \geq -1/2$ and $\Delta_{\alpha,\beta} = x^\beta D_x x^{\alpha-\beta+1}D_x x^{-\alpha} = xD_x^2 - (\alpha+\beta-1)D_x + \alpha\beta x^{-1}$ denotes the generalized Kepinski operator. $\mathbb{H}_{\alpha,\beta,a}$ is a Frechet space. In [2] we established that the kernel of (28) belongs to $\mathbb{H}_{\alpha,\beta,a}$ and defined the distributional generalized Hankel–Clifford transformation $F'_{\alpha,\beta}$ on the dual space $\mathbb{H}'_{\alpha,\beta,a}$ by the relation

$$\begin{aligned} F'_{\alpha,\beta}\{f\}(y) = F_1(y) &= \left\langle f(x), \left(\frac{y}{x}\right)^{-(\alpha+\beta)/2} J_{\alpha-\beta}(2\sqrt{xy})\right\rangle\\ &= \langle f(x), y^{-(\alpha+\beta)}C_{\alpha,\beta}(xy)\rangle \end{aligned} \tag{48}$$

where $f \in \mathbb{H}'_{\alpha,\beta,a}(I)$.

Lemma 5.2. *Let* $(\alpha-\beta)\geq -1/2$. $\mathbb{S}_\alpha(I)$ *is a subspace of* $\mathbb{H}_{\alpha,\beta,a}$, *the topology of* $\mathbb{S}_\alpha(I)$ *being stronger than that induced on it by* $\mathbb{H}_{\alpha,\beta,a}$. *Consequently restriction of* $f\in\mathbb{H}'_{\alpha,\beta,a}$ *to* $\mathbb{S}_\alpha(I)$ *is in* $\mathbb{S}'_\alpha(I)$.

Proof. Proof can be easily given [2].

Note that $(y/x)^{-(\alpha+\beta)/2}J_{\alpha-\beta}(2\sqrt{xy})$ belongs to $\mathbb{H}_{\alpha,\beta,a}$, [2] for all y fixed in $0<y<\infty$, but not to $\mathbb{S}_\alpha(I)$, because this function is not of rapid descent at infinity. Hence the above inclusion is proper.

Every member of $f\in\mathbb{H}'_{\alpha,\beta,a}$ admits, by Lemma 5.2, two distributional generalized Hankel–Clifford transformations given through (44) and (48). Our next objective is to show that these definitions agree.

Theorem 4. *Let* $(\alpha-\beta)\geq -1/2$. *If* $f\in\mathbb{H}'_{\alpha,\beta,a}$ *then the distributional generalized Hankel–Clifford transformation* $F'_{\alpha,\beta}\{f\}(y)$ *defined by* (48) *coincides with* $h'_{1,\alpha,\beta}$ *given by* (44), *in the sense of equality in* $\mathbb{S}'_\alpha(I)$.

Proof. Proof can be easily given [2].

Remark 2. If the distributional generalized Hankel–Clifford transformation $h'_{1,\alpha,\beta}$ were defined on $\mathbb{H}'_\beta(I)$, as usual in the available literature, by means of the adjoint of $h_{1,\alpha,\beta}$ on $\mathbb{H}_\beta(I)$, namely

$$\langle h'_{1,\alpha,\beta}f,\Phi\rangle=\langle f,h_{1,\alpha,\beta}\Phi\rangle,\quad f\in\mathbb{H}'_\beta(I),\quad \Phi\in\mathbb{H}_\beta(I)\tag{49}$$

the relation (46) between the classical and the distributional transforms must be replaced by

$$h'_{1,\alpha,\beta}f=y^{-(\alpha+\beta)}h_{1,\alpha,\beta}(x^{(\alpha+\beta)}f).$$

In other words, the classical transform $h_{1,\alpha,\beta}$ is not a special case of the distributional transform $h'_{1,\alpha,\beta}$. On the other hand, the operational formulas of $h_{1,\alpha,\beta}$ and $h'_{1,\alpha,\beta}$ would not coincide. Another drawback of this usual definition lies in the fact that Lemma 5.1 could not be established. Moreover (49) can not be understood as a generalization of Parseval equation (14) due to presence of the factor $x^{(\alpha+\beta)}$. These considerations show that our definition (44) is more adequate than (49).

Remark 3. Analogously, the distributional generalized Hankel–Clifford transformation $h'_{2,\alpha,\beta}$ can be defined on $\mathbb{H}'_\beta(I)$ as the adjoint of $h_{1,\alpha,\beta}$ on $\mathbb{H}_\beta(I)$, that is,

$$\langle h'_{2,\alpha,\beta}f,\Phi\rangle=\langle f,h_{1,\alpha,\beta}\Phi\rangle,\quad f\in\mathbb{H}'_\beta(I),\quad \Phi\in\mathbb{H}_\beta(I).$$

$h'_{2,\alpha,\beta}$ is an automorphism too on $\mathbb{H}'_\beta(I)$ by virtue of Theorem 2, provided that $(\alpha-\beta)\geq -1/2$. Since $\mathbb{S}_\alpha(I)\subset\mathbb{H}'_\beta(I)$ and $\mathbb{S}_\alpha(I)\subset\mathbb{H}'_{\beta-1}(I)$, the operational formulas for $h'_{2,\alpha,\beta}$ turn out to be those in Lemma 4.2.

6. The distributional generalized Hankel–Clifford transformation of arbitrary order

Let $(\alpha-\beta)$ be any fixed real number. We first define a certain transformation on $\mathbb{H}_\beta(I)$, which coincides with (4) whenever $(\alpha-\beta)\geq -1/2$. Let k be any positive integer such that $(\alpha-\beta+k)\geq -1/2$. For any $\Phi\in \mathbb{H}_\beta(I)$ set

$$\phi(x)=h_{1,\alpha,\beta,k}[\Phi(y)]=(-1)^k x^{-k}h_{1,\alpha,\beta-k}R_{\beta-(k-1)}\cdots R_{\beta-1}R_\beta\Phi(y) \qquad (5$$

$$\Phi(y)=h^{-1}_{1,\alpha,\beta,k}[\phi(x)]=(-1)^k R^{-1}_\beta R^{-1}_{\beta-1}\cdots R^{-1}_{\beta-(k-1)}R^{-1}_{\beta-1}h^{-1}_{1,\alpha,\beta-k}x^k\phi(x). \qquad (5$$

Lemma 6.1. *The transformation $h_{1,\alpha,\beta,k}$ as defined by* (50) *is an automorphism on $\mathbb{H}_\beta($ whatever be the real number $(\alpha-\beta)$. Its inverse is $h^{-1}_{1,\alpha,\beta,k}$ as defined by* (51). *Finally, $\mathbb{H}_\beta(I)$, $h_{1,\alpha,\beta,k}$ coincides with $h_{1,\alpha,\beta}$ as defined by* (6) *whenever $(\alpha-\beta)\geq -1/2$.*

Proof. The first assertion follows from the fact that $\Phi\rightarrow R_{\beta-(k-1)}\cdots R_{\beta-1}R_\beta\Phi$ is isomorphism from $\mathbb{H}_\beta(I)$ onto $\mathbb{H}_{\beta-k}(I)$, $\Phi\rightarrow h_{1,\alpha,\beta,k}\Phi$ is an automorphism on $\mathbb{H}_{\beta-k}($ and $\phi\rightarrow x^{-k}\phi$ is an isomorphism from $\mathbb{H}_{\beta-k}(I)$ onto $\mathbb{H}_\beta(I)$.

By assumption, $\alpha+\beta+k\geq -1/2$. It is a classical fact that $h_{1,\alpha,\beta-k}$ is the inverse itself when it acts on smooth functions in $L(I)$. Since $\mathbb{H}_{\beta-k}(I)\subset L(I)$, the seco assertion follows from this fact and Lemma 2.1, (ii), Lemma 2.2.

For the third assertion, assume that $\Phi(y)\in\mathbb{H}_\beta(I)$ and $(\alpha-\beta)\geq -1/2$, and consi the case $k=1$;

$$\begin{aligned}h_{1,\alpha,\beta,1}\Phi&=-x^{-1}h_{1,\alpha,\beta-1}\Phi\\&=-x^{-1}x^{-(\alpha+\beta-1)}\int_0^\infty C_{\alpha,\beta-1}(xy)(y^{-\beta+1}D_y y^\beta\Phi(y))\mathrm{d}y\end{aligned}$$

An integration by parts and the formula

$$D_y[y^{-\beta+1}C_{\alpha,\beta-1}(xy)]=y^{-\beta}C_{\alpha,\beta}(xy)$$

yield

$$=-x^{-1}x^{-(\alpha+\beta-1)}[y^{-\beta+1}C_{\alpha,\beta-1}(xy)(y^{-\beta}\Phi(y)]_0^\infty-\int_0^\infty y^{-\beta}C_{\alpha,\beta}(xy)y^\beta\Phi(y)\mathrm{c}$$

The limit terms are zero because $\Phi(y)$ is of rapid descent and $yC_{\alpha,\beta-1}(xy)$ rema bounded as $y\rightarrow\infty$, whereas, for $y\rightarrow 0^+$, $C_{\alpha,\beta}(x)=0(x^\alpha)$ where $(\alpha-\beta)\geq -1/2$. Th

$$h_{1,\alpha,\beta,1}(\Phi)=x^{-1}x^{-(\alpha+\beta-1)}\int_0^\infty C_{\alpha,\beta}(xy)\Phi(y)\mathrm{d}y=h_{1,\alpha,\beta}(\Phi).$$

The general statement for larger integral values of k follows by induction from results. This ends the proof.

A consequences of Lemma 6.1 is that $h_{1,\alpha,\beta,k}=h_{1,\alpha,\beta,p}$ so long as the positive inte k and p are both larger than $-(\alpha-\beta)-1/2$. Indeed, assuming that $k>p$, we h $h_{1,\alpha,\beta-p,k+p}=h_{1,\alpha,\beta-p}$ according to the last statement of Lemma 6.1, and therefore, $\Phi\in\mathbb{H}_\beta(I)$,

$$h_{1,\alpha,\beta,k}\Phi=(-1)^p x^{-p}h_{1,\alpha,\beta-p,k+p}R_{\beta-(p-1)}\cdots R_\beta\Phi=h_{1,\alpha,\beta,p}\Phi.$$

Since $h_{1,\alpha,\beta}\Phi=h^{-1}_{1,\alpha,\beta}\Phi$ whenever $\Phi\in\mathbb{H}_\beta(I)$ and $(\alpha-\beta)\geq -1/2$. Lemma 6.1 implies that $h^{-1}_{1,\alpha,\beta,k}$ coincides with $h^{-1}_{1,\alpha,\beta}$ when $(\alpha-\beta)\geq -1/2$. Moreover, by vi of the preceding paragraph, $h^{-1}_{1,\alpha,\beta,k}$ is independent of the choice of k so long $\alpha-\beta+k\geq -1/2$.

If view of these results, it is reasonable to define the first generalized Hankel–Clif transformation $h_{1,\alpha,\beta,k}$ for $(\alpha-\beta)<-1/2$ on $\Phi\in\mathbb{H}_\beta(I)$ by $h_{1,\alpha,\beta}\Phi=h_{1,\alpha,\beta,k}\Phi$ whe is any positive integer no less than $-(\alpha-\beta)-1/2$. The generalized inverse Hankel–Clifford transformation $h^{-1}_{1,\alpha,\beta}$ is defined by $h^{-1}_{1,\alpha,\beta}\Phi = h^{-1}_{1,\alpha,\beta,k}\Phi$, $\Phi \in \mathbb{H}_\beta(I)$. As in the classical case, $h_{1,\alpha,\beta} = h^{-1}_{1,\alpha,\beta}$ when $(\alpha-\beta) \geq -1/2$, but this is not the case when $(\alpha-\beta) < -1/2$.

Lemma 6.2. *Let* $(\alpha-\beta)$ *be any real number. For all* $\phi(x) \in \mathbb{H}_\beta(I)$, *we have*

$$\begin{aligned} R_\beta h_{1,\alpha,\beta,k}(\phi) &= h_{1,\alpha,\beta-1,k}(-x\phi),\\ h_{1,\alpha,\beta-1,k}(R_\beta\phi) &= -y\, h_{1,\alpha,\beta,k}(\phi),\\ h_{1,\alpha,\beta,k}(Q_\alpha R_\beta\phi) &= -y\, h_{1,\alpha,\beta,k}(\phi),\\ Q_\alpha R_\beta h_{1,\alpha,\beta,k}(\phi) &= h_{1,\alpha,\beta,k}(-x\phi). \end{aligned}$$

and for all $\phi(x) \in \mathbb{H}_{\beta-1}(I)$

$$\begin{aligned} h_{1,\alpha,\beta,k}(Q_\alpha\phi) &= h_{1,\alpha,\beta-1,k}(\phi),\\ Q_\alpha h_{1,\alpha,\beta-1,k}(\phi) &= h_{1,\alpha,\beta,k}(\phi). \end{aligned}$$

Now we define the second generalized Hankel–Clifford transformation of arbitrary order for $h_{2,\alpha,\beta}$.

Let $\alpha-\beta$ be any fixed real number and k any positive integer such that $\alpha-\beta+k \geq -1/2$. We define the transformation $h_{2,\alpha,\beta,k}$ on any $\psi \in \mathbb{S}_\alpha(I)$ by

$$h_{2,\alpha,\beta,k}[\Psi(y)] = \Psi(x) = h_{2,\alpha,\beta-k},(Q^*_\alpha)^k\Psi(y) \tag{52}$$

and

$$\psi(y) = h^{-1}_{2,\alpha,\beta,k}[\psi(x)] = \{(Q^*_\alpha)^{-1}\}^k h_{2,\alpha,\beta-k}\Psi(x). \tag{53}$$

Analogous conclusions as Lemma 6.1 is also valid.

Lemma 6.3. *The transformation* $h_{2,\alpha,\beta,k}$ *as defined by* (52) *is an automorphism on* $\mathbb{S}_\alpha(I)$ *whatever be the real number* $\alpha-\beta$. *Its inverse is* $h^{-1}_{2,\alpha,\beta,k}$ *as defined by* (53). *Finally, on* $\mathbb{S}_\alpha(I)$, $h_{2,\alpha\beta,k}$ *coincides with* $h_{2,\alpha,\beta}$ *as defined by* (5) *whenever* $\alpha-\beta \geq -1/2$.

We now turn to the distributional generalized Hankel–Clifford transformation $h'_{1,\alpha,\beta}$ of any arbitrary order $(\alpha-\beta)$ when it is acting on $\mathbb{S}'_\alpha(I)$. As before, k is any positive integer $\geq -(\alpha-\beta)-1/2$. Then the distributional generalized Hankel–Clifford transformation $h'_{1,\alpha,\beta}$ is defined on $\mathbb{S}'_\alpha(I)$ as the adjoint of $h_{2,\alpha,\beta,k}$ on $\mathbb{S}_\alpha(I)$ as

$$\langle h'_{1,\alpha,\beta}f,\Phi\rangle = \langle f, h_{2,\alpha,\beta,k}\Phi\rangle, \quad \text{for } f \in \mathbb{S}'_\alpha(I), \quad \Phi \in \mathbb{S}_\alpha(I). \tag{54}$$

Theorem 5. *The distributional Hankel–Clifford transformation of arbitrary order* $(\alpha-\beta)$, *defined in* (54), *is an automorphism onto* $\mathbb{S}'_\alpha(I)$.

This leads to the following transformation formulas:

Lemma 6.4. *Let* $(\alpha-\beta)$ *be any real number. For* $f \in \mathbb{S}'_\alpha(I)_\alpha$,

$$\begin{aligned} R_\beta h'_{1,\alpha,\beta}(\phi) &= h'_{1,\alpha,\beta-1}(-x\phi),\\ h'_{1,\alpha,\beta-1}(R_\beta\phi) &= -y h'_{1,\alpha,\beta}(\phi), \end{aligned}$$

$$h'_{1,\alpha,\beta}(Q_\alpha R_\beta \phi) = -y h'_{1,\alpha,\beta}(\phi),$$
$$Q_\alpha R_\beta h'_{1,\alpha,\beta}(\phi) = h'_{1,\alpha,\beta}(-x\phi).$$

and for all $\phi(x) \in \mathbb{S}_{\alpha+1}(I)$

$$h'_{1,\alpha,\beta}(Q_\alpha \phi) = h'_{1,\alpha,\beta-1}(\phi),$$
$$Q_\alpha h'_{1,\alpha,\beta-1}(\phi) = h'_{1,\alpha,\beta}(\phi).$$

Note. Analogously the distributional generalized Hankel–Clifford transformation $h'_{2,\alpha,\beta}$ can be defined on $\mathbb{H}'_\beta(I)$ as the adjoint of $h_{1,\alpha,\beta,k}$ on $\mathbb{H}_\beta(I)$ as $\langle h'_{2,\alpha,\beta} f, \Psi\rangle = \langle f, h_{1,\alpha,\beta,k}\Psi\rangle$ for $f \in \mathbb{H}'_\beta(I)$ and $\Psi \in \mathbb{H}_\beta(I)$.

Remark 4. It is proposed to apply the theory thus developed to solve some partial differential equations of generalized Kapinski type operator with distributional boundary conditions in subsequent papers.

References

[1] Koh E L and Zemanian A H, The complex Hankel and I transformations of generalized functions, *SIAM J. Appl. Math.* **16(5)** (1968) 945–957

[2] Malgonde S P, Generalized Hankel–Clifford transformation on certain spaces of distributions, communicated for publication

[3] Mendez J M R Perez and Socas Robayana M, A pair of generalized Hankel–Clifford transformations and their applications. *J. Math. Anal. Appl.* **154** (1991) 543–557

[4] Mendez Perez J M R and Socas Robayana M, La Transformacion Integral De Hankel–Clifford De Orden Arbitrario Homenaje al Prof. Dr. Nacere Hayek Calil (1990) pp. 199–207

[5] Zemanian A H, Hankel transforms of arbitrary order, *Duke Math. J.* **34** (1967) 761–767

[6] Zemanian A H, Generalized integral transformation, *Interscience* (1966) (republished by Dover, New York, 1987)

Proc. Indian Acad. Sci. (Math. Sci.), Vol. 110, No. 3, August 2000, pp. 305–314.
© Printed in India

C^2-rational cubic spline involving tension parameters

M SHRIVASTAVA and J JOSEPH

Department of Mathematics and Computer Science, R D University, Jabalpur 482 001, India

MS received 8 September 1998; revised 15 January 2000

Abstract. In the present paper, C^1-piecewise rational cubic spline function involving tension parameters is considered which produces a monotonic interpolant to a given monotonic data set. It is observed that under certain conditions the interpolant preserves the convexity property of the data set. The existence and uniqueness of a C^2-rational cubic spline interpolant are established. The error analysis of the spline interpolant is also given.

Keywords. Interpolation; rational spline; tension parameter; monotonicity; convexity; continuity.

1. Introduction

Interpolation techniques play a very important role in obtaining solutions of various problems that arise in many areas of scientific computation. Generally interpolant is preferred which preserves some of the characteristics of the function to be interpolated. In order to tackle such situations, a variety of shape preserving interpolation methods have been discussed in the literature for interpolating given sets of monotonic and convex data.

Shape preserving cubic spline interpolants have been obtained by Fritsch and Carlson in [4]. Gregory and Delbourgo, being motivated by the work of Fritsch and Carlson have introduced piecewise rational interpolatory splines (cf. [5]). Using rational functions, problem of obtaining a convex interpolant to convex data has been investigated by Delbourgo (cf. [3]). Shape preserving properties of the rational (cubic/quadratic) spline interpolant have been studied in [2]. In order to control the shape of the rational interpolant in a more efficient manner, Delbourgo and Gregory (cf. [2]) have introduced tension parameters in the definition of the C^1-rational cubic spline function. The tension parameters have been so chosen that it provides the desired geometric shape to the rational interpolant. For a designer these tension parameters act as intuitive tools for manipulating the shape of the curve. A C^2-piecewise rational (cubic/cubic) Bézier curve involving two tension parameters which is used to interpolate the given monotonic data is described in [6]. Shape preserving (cubic/linear) rational interpolants to monotone and convex functions have been studied in [7].

In the present paper, a rational spline interpolant is constructed which matches given data values and at the same time preserves certain geometric features namely the monotonicity and convexity properties of the function to be interpolated. In fact, we obtain a rational (cubic/linear) spline interpolant involving two tension parameters when values of the function and its first derivative are given at the knots.

Introducing two parameters r_i and t_i, we construct a C^1-rational (cubic/linear) spline interpolant and obtain its error bounds in § 2. In § 3, we study the convexity and monotonicity of this rational spline interpolant. Existence of a unique C^2-rational spline interpolant has been discussed in § 4. In § 5, we consider a numerical example and discuss the impact of variation of parameters r_i and t_i on the shape of the interpolant. Some remarks are given in § 6.

2. The rational spline interpolant

Let $P = \{x_i\}_{i=1}^{n}$ where $a = x_1 < x_2 < \cdots < x_n = b$, be a partition of the interval $[a, b]$, let f_i, $i = 1, \ldots, n$ be the function values at the data points. We set

$$h_i = x_{i+1} - x_i, \quad \Delta_i = (f_{i+1} - f_i)/h_i \tag{2.1}$$

and

$$\theta = (x - x_i)/h_i. \tag{2.2}$$

Further, we set

$$s(x) = P_i(\theta)/Q_i(\theta), \tag{2.3}$$

where

$$\begin{aligned} P_i(\theta) =& r_i f_i (1-\theta)^3 + \{(2r_i + t_i) f_i + r_i h_i d_i\}\theta(1-\theta)^2 \\ &+ \{(r_i + 2t_i) f_{i+1} - t_i h_i d_{i+1}\}\theta^2(1-\theta) + t_i f_{i+1}\theta^3 \end{aligned} \tag{2.4}$$

and

$$Q_i(\theta) = r_i + (t_i - r_i)\theta. \tag{2.5}$$

Here we choose parameters r_i and t_i in such a manner that

$$r_i, t_i > 0 \quad \text{and} \quad t_i > r_i \tag{2.6}$$

which ensures a strictly positive denominator in the rational spline. It is easily seen that

$$\begin{aligned} s^{(1)}(x) =& \{[r_i + (t_i - r_i)\theta]\{-3r_i f_i (1-\theta)^2 + \{(2r_i + t_i) f_i \\ &+ r_i h_i d_i\}[(1-\theta)^2 - 2\theta(1-\theta)] + \{(r_i + 2t_i) f_{i+1} \\ &- t_i h_i d_{i+1}\}[2\theta(1-\theta) - \theta^2] + 3t_i f_{i+1}\theta^2\} \\ &+ (r_i - t_i)[r_i f_i (1-\theta)^3 + \{(2r_i + t_i) f_i + r_i h_i d_i\}\theta(1-\theta)^2 \\ &+ \{(r_i + 2t_i) f_{i+1} - t_i h_i d_{i+1}\}\theta^2(1-\theta) \\ &+ t_i f_{i+1}\theta^3]\}/h_i[r_i + (t_i - r_i)\theta]^2. \end{aligned} \tag{2.7}$$

We observe that

$$\begin{aligned} &s(x_i) = f_i, \quad && s(x_{i+1}) = f_{i+1}, \\ &s^{(1)}(x_i) = d_i, \quad && s^{(1)}(x_{i+1}) = d_{i+1}, \end{aligned} \tag{2.8}$$

where d_i's denote the derivative values at the knots x_i. These derivative parameters are usually not given and can be determined by using the methods as discussed in [1].

Let $e = s - f$ denote the error function. We prove the following theorem.

Theorem 2.1. *Given a function f, let s be its piecewise rational (cubic/linear) spline interpolant satisfying the interpolatory conditions (2.8). Then for $x \in [x_i, x_{i+1}]$, the following hold.*

(a) *If $f \in C^4[a,b]$, then*

$$\| e \| \leq \frac{h_i}{4\underline{r}} \max\{\bar{r}|f_i^{(1)} - d_i|, \bar{t}|f_{i+1}^{(1)} - d_{i+1}|\} + \frac{\bar{t}}{384\underline{r}} \{h_i^4 \| f^{(4)} \| + 4h_i^3 \| f^{(3)} \| .$$

(b) *If $f \in C^1[a,b]$, then*

$$\| e \| \leq \frac{13}{9\underline{r}}(\bar{r} + \bar{t})\omega(f,h) + \frac{\bar{t}}{4\underline{r}} h_i \max\{|d_i|, |d_{i+1}|\},$$

where $\bar{r} = \max r_i$, $\bar{t} = \max t_i$, $\underline{r} = \min r_i$ and $\| \cdot \|$ denotes the uniform norm on $[x_i, x_{i+1}]$.

Proof. We set $x(\theta) = (x_i + \theta h_i)$ and $f(x) = F_i(\theta), 0 \leq \theta \leq 1$ in $[x_i, x_{i+1}]$. Then we observe that, in $[x_i, x_{i+1}]$,

$$e(x) = f(x) - s(x) = F_i(\theta) - P_i(\theta)/Q_i(\theta). \tag{2.9}$$

Let us take

$$\begin{aligned} R_i(\theta) = {} & r_i f_i (1-\theta)^3 + [(2r_i + t_i) f_i + r_i h_i f_i^{(1)}]\theta(1-\theta)^2 \\ & + [(r_i + 2t_i) f_{i+1} - t_i h_i f_{i+1}^{(1)}]\theta^2(1-\theta) + t_i f_{i+1}\theta^3, \end{aligned} \tag{2.10}$$

and

$$s_i(\theta) = F_i(\theta) Q_i(\theta) = f(x_i + \theta h_i)[r_i + (t_i - r_i)\theta]. \tag{2.11}$$

We observe that

$$\begin{aligned} |F_i(\theta) - P_i(\theta)/Q_i(\theta)| & \leq [|F_i(\theta)Q_i(\theta) - R_i(\theta)| + |R_i(\theta) - P_i(\theta)|]/|Q_i(\theta)| \\ & = [|s_i(\theta) - R_i(\theta)| + |R_i(\theta) - P_i(\theta)|]/|Q_i(\theta)|. \end{aligned} \tag{2.12}$$

It can be verified that

$$\begin{aligned} & R_i(0) = s_i(0) = r_i f_i; & & R_i^{(1)}(0) = s_i^{(1)}(0) = (t_i - r_i) f_i + r_i h_i f_i^{(1)}; \\ & R_i(1) = s_i(1) = t_i f_{i+1}; & & R_i^{(1)}(1) = s_i^{(1)}(1) = (t_i - r_i) f_{i+1} + t_i h_i f_{i+1}^{(1)}. \end{aligned}$$

Thus $R_i(\theta)$ is the cubic Hermite interpolant to $s_i(\theta)$ on $0 \leq \theta \leq 1$. Now bounding Cauchy remainder of $s_i(\theta)$, we get

$$\begin{aligned} |s_i(\theta) - R_i(\theta)| & \leq \frac{1}{384} \max_{0 \leq \theta \leq 1} \left| \frac{\mathrm{d}^4}{\mathrm{d}\theta^4} s_i(\theta) \right| \\ & = \frac{1}{384} \max_{0 \leq \theta \leq 1} |F_i^{(4)}(\theta) Q_i(\theta) + 4F_i^{(3)}(\theta) Q_i^{(1)}(\theta)|, \end{aligned}$$

since $Q_i(\theta)$ is linear, so that $Q_i^{(2)}(\theta) = Q_i^{(3)}(\theta) = 0$. Now

$$|Q_i(\theta)| = |r_i + (t_i - r_i)\theta| \leq \max_i \{t_i\} = \bar{t},$$

$$|Q_i^{(1)}(\theta)| = |t_i - r_i| \leq \max_i (t_i - r_i) \leq \bar{t},$$

and

$$|F_i^{(j)}(\theta)| \le h_i^j \parallel f^{(j)} \parallel .$$

Hence

$$|s_i(\theta) - R_i(\theta)| \le \frac{\bar{t}}{384}\{h_i^4 \parallel f^{(4)} \parallel + 4h_i^3 \parallel f^{(3)} \parallel\}. \qquad (2.1$$

Also

$$\begin{aligned} |R_i(\theta) - P_i(\theta)| &\le h_i\theta(1-\theta)\{|r_i(1-\theta)(f_i^{(1)} - d_i)| + |t_i\theta(d_{i+1} - f_{i+1}^{(1)}|\} \\ &\le \frac{h_i}{4}\max\{\bar{r}|f_i^{(1)} - d_i|, \bar{t}|f_{i+1}^{(1)} - d_{i+1}|\}. \qquad (2.1 \end{aligned}$$

Considering the denominator in (2.3) we find that

$$\min|Q_i(\theta)| = \min\{|r_i + (t_i - r_i)\theta|\} > \underline{r}. \qquad (2.1$$

Combining (2.13), (2.14) and (2.15) in (2.12) and (2.9) we get

$$\begin{aligned} \parallel e \parallel \le{} & \frac{h_i}{4\underline{r}}\max\{\bar{r}|f_i^{(1)} - d_i|, \bar{t}|f_{i+1}^{(1)} - d_{i+1}|\} \\ & + \frac{\bar{t}}{384\underline{r}}\{h_i^4 \parallel f^{(4)} \parallel + 4h_i^3 \parallel f^{(3)} \parallel\}. \end{aligned}$$

This completes the proof of (a).

Further, from (2.3)–(2.5) we observe that

$$s(x) - f(x) = [P_i(\theta) - (r_i + (t_i - r_i)\theta)f(x)]/[r_i + (t_i - r_i)\theta],$$

so that

$$\begin{aligned} |s(x) - f(x)| \le{} & [|r_i(1-\theta)^3 + (2r_i + t_i)\theta(1-\theta)^2||f_i - f(x)| \\ & + |(2t_i + r_i)\theta^2(1-\theta) + t_i\theta^3||f_{i+1} - f(x)| \\ & + h_i\theta(1-\theta)|r_i d_i(1-\theta) - t_i d_{i+1}\theta|]/\min\{|r_i + (t_i - r_i)\theta|\} \\ \le{} & \frac{13}{9\underline{r}}(\bar{r} + \bar{t})\omega(f,h) + \frac{\bar{t}}{4\underline{r}}h_i\max\{|d_i|, |d_{i+1}|\}. \end{aligned}$$

Thus

$$\|e\| \le \frac{13}{9\underline{r}}(\bar{r} + \bar{t})\omega(f,h) + \frac{\bar{t}}{4\underline{r}}h_i\max\{|d_i|, |d_{i+1}|\}, \qquad (2.$$

which proves (b).

This completes the proof of Theorem 2.1.

3. Convexity and monotonicity of the interpolant

We shall now investigate the monotonicity and convexity preserving properties of rational (cubic/linear) interpolant to a given monotonic or convex data.

3.1. *Monotonicity*

Let f be a monotonic increasing function in $[a,b]$ so that

$$f_1 \leq f_2 \leq \cdots \leq f_n, \text{ or equivalently } \Delta_i \geq 0. \tag{3.1}$$

We choose the derivative values d_i such that

$$d_i \geq 0, \quad i = 1, \ldots, n. \tag{3.2}$$

We observe that $s(x)$ is monotonic increasing if and only if for $x \in [a,b]$,

$$s^{(1)}(x) \geq 0. \tag{3.3}$$

A simple manipulation in (2.7) shows that for $x \in [x_i, x_{i+1}]$,

$$\begin{aligned} s^{(1)}(x) = {} & [r_i^2 d_i(1-\theta)^3 + X_i\theta(1-\theta)^2 + Y_i\theta^2(1-\theta) \\ & + t_i^2 d_{i+1}\theta^3]/[r_i + (t_i - r_i)\theta]^2, \end{aligned} \tag{3.4}$$

where

$$X_i = 2r_i^2\Delta_i + 4r_it_i\Delta_i - r_i^2 d_i - 2r_it_id_{i+1}$$

and

$$Y_i = 2t_i^2\Delta_i - 2r_it_id_i - t_i^2 d_{i+1} + 4r_it_i\Delta_i.$$

We observe that the denominator of rational function $s^1(x)$ given in (2.7) is positive. Therefore considering the numerator in (3.4), we find that $s^{(1)}(x)$ is non-negative if $X_i \geq 0, Y_i \geq 0$ provided (3.2) is also satisfied. We observe that the sufficient condition that $X_i \geq 0$ and $Y_i \geq 0$ is

$$\frac{r_i}{t_i} \geq \frac{(d_{i+1} - \Delta_i)}{(\Delta_i - d_i)}. \tag{3.5}$$

Therefore $s^{(1)}(x)$ is non-negative if (3.5) holds.

We have thus proved the following theorem:

Theorem 3.1. *Given a monotonic increasing set of data satisfying* (3.1) *and the derivative values satisfying* (3.2), *there exists a monotone rational* (*cubic/linear*) *spline interpolant* $s \in C^1[a,b]$ *involving the tension parameters* r_i *and* t_i *which satisfies the interpolatory conditions* (2.8) *provided* (3.5) *holds.*

3.2. *Convexity*

Suppose the given data set is strictly convex then

$$\Delta_1 < \Delta_2 < \cdots < \Delta_{n-1}.$$

We choose derivative values $d_i \geq 0$ to be such that

$$0 \leq d_1 < \Delta_1 < d_2 < \cdots < \Delta_{i-1} < d_i < \Delta_i < \cdots < d_n. \tag{3.6}$$

A simple calculation shows that for $x \in [x_i, x_{i+1}]$, from (2.7) we get

$$\begin{aligned} s^{(2)}(x) = {} & [A_{2i}(1-\theta)^3 + B_{2i}\theta(1-\theta)^2 + C_{2i}\theta^2(1-\theta) \\ & + D_{2i}\theta^3]/h_i[r_i + (t_i - r_i)\theta]^3, \end{aligned} \tag{3.7}$$

where

$$\begin{aligned}
A_{2i} &= 2r_i^3\Delta_i + 4r_i^2t_i\Delta_i - 2r_i^3d_i - 2r_i^2t_id_{i+1} - 2r_i^2t_id_i,\\
B_{2i} &= 6r_i^2t_i(\Delta_i - d_i),\\
C_{2i} &= 6r_it_i^2(d_{i+1} - \Delta_i) \text{ and}\\
D_{2i} &= 2t_i^3(d_{i+1} - \Delta_i) + 2r_it_i^2(d_{i+1} - \Delta_i) + 2r_it_i^2(-\Delta_i + d_i).
\end{aligned} \tag{3.8}$$

We observe that $s^{(2)}(x)$ is non-negative if each of A_{2i}, B_{2i}, C_{2i} and D_{2i} is non-negative. Since we are assuming that (3.6) holds, B_{2i} and C_{2i} are automatically positive. Thus the sufficient condition for the interpolant $s(x)$ to be convex is that $A_{2i} \geq 0$ and $D_{2i} \geq 0$.

Now $A_{2i} \geq 0$ if $2r_i^3(\Delta_i - d_i) - 2r_i^2t_i(d_{i+1} - \Delta_i) + 2r_i^2t_i(\Delta_i - d_i) \geq 0$,

$$\text{i.e., if } \frac{r_i}{t_i} \geq \frac{(d_{i+1} - \Delta_i)}{(\Delta_i - d_i)} \tag{A}$$

and $D_{2i} \geq 0$ if $2t_i^3(d_{i+1} - \Delta_i) + 2r_it_i^2(-\Delta_i + d_i) \geq 0$,

$$\text{i.e., if } \frac{r_i}{t_i} \leq \frac{(d_{i+1} - \Delta_i)}{(\Delta_i - d_i)}. \tag{B}$$

Therefore the spline interpolant is convex if

$$\frac{r_i}{t_i} = \frac{(d_{i+1} - \Delta_i)}{(\Delta_i - d_i)}, \tag{3.9}$$

provided (2.6) and (3.6) hold.

Thus the spline interpolant is convex if (3.9) together with (2.6) and (3.6) holds.

We have thus proved the following theorem.

Theorem 3.2. *For a given set of strictly convex data, a convex rational* (*cubic/linear*) *spline interpolant $s \in C^1[a, b]$ involving the parameters r_i and t_i exists which satisfies the interpolatory conditions* (2.8), *with the derivative parameters d_i's satisfying* (3.6) *provided* (2.6) *and* (3.9) *hold.*

4. C^2-rational spline interpolant

For a given set of data points $\{(x_i, f_i)\}_{i=1}^n$, let s defined in § 2, represent a C^2-rational (cubic/linear) spline interpolant.

For $x \in [x_i, x_{i+1}]$, we have

$$\begin{aligned}
s^{(2)}(x) = {} & [A_{2i}(1-\theta)^3 + B_{2i}\theta(1-\theta)^2 + C_{2i}\theta^2(1-\theta)\\
& + D_{2i}\theta^3]/h_i[r_i + (t_i - r_i)\theta]^3,
\end{aligned}$$

where A_{2i}, B_{2i}, C_{2i} and D_{2i} are given by (3.8). It is easy to see that

$$\begin{aligned}
s^{(2)}(x_{i-}) &= D_{2i-1}/h_{i-1}t_{i-1}^3\\
&= [2r_{i-1}t_{i-1}^2(d_i - 2\Delta_{i-1} + d_{i-1}) + 2t_{i-1}^3(d_i - \Delta_{i-1})]/h_{i-1}t_{i-1}^3\\
&= 2[r_{i-1}d_{i-1} - (2r_{i-1} + t_{i-1})\Delta_{i-1} + (r_{i-1} + t_{i-1})d_i]/h_{i-1}t_{i-1}
\end{aligned}$$

and

$$\begin{aligned}s^{(2)}(x_{i+}) &= A_{2i}/h_i r_i^3\\ &= 2[(r_i+2t_i)\Delta_i - t_i d_{i+1} - (r_i+t_i)d_i]/h_i r_i.\end{aligned}$$

Therefore continuity of $s^{(2)}$ gives that

$$\begin{aligned}&h_i r_i r_{i-1} d_{i-1}[h_{i-1}t_{i-1}(r_i+t_i) + h_i r_i(r_{i-1}+t_{i-1})]d_i\\ &\quad + h_{i-1}t_i t_{i-1} d_{i+1} = h_i r_i(2r_{i-1}+t_{i-1})\Delta_{i-1} + h_{i-1}t_{i-1}(r_i+2t_i)\Delta_i,\end{aligned} \tag{4.1}$$

where the Δ_i's, $i = 1, \ldots, n$ are given by (2.1).

We observe that if d_1 and d_n are the given quantities then (4.1) represents a system of $(n-2)$ equations in $(n-2)$ unknowns, namely $d_2, \ldots, d_{n-1}$. Assume that $r_i \geq r > 0$ and $t_i \geq t > 0$, $i = 1, \ldots, n-1$, then it is easy to see that the coefficients of d_{i-1}, d_i and d_{i+1} are all positive and the excess of coefficient of d_i over the sum of those of d_{i-1} and d_{i+1} is $(h_{i-1} + h_i)r_i t_{i-1}$ which is clearly positive. Therefore the coefficient matrix of the system of equations (4.1) is diagonally dominant and is thus invertible. Therefore a unique solution for the system of equations (4.1) exists.

This establishes the following theorem.

Theorem 4.1. *Let f_i, $i = 1, \ldots, n$ be the given data-values. Then for the derivative parameters given at the end points namely d_i and d_n, there exists a unique C^2-piecewise rational (cubic/linear) spline interpolant satisfying the interpolatory conditions* (2.8) *provided that* (2.6) *holds.*

5. Numerical example

In this section, we construct a C^2-rational (cubic/linear) spline interpolant which involves two tension parameters r_i and t_i for a given set of function values. In view of (4.1), suppose that the function values f_i's at the knots and the derivative values at the end points namely d_1 and d_n are given. The derivative parameters d_i's, $i = 2, \ldots, n-1$ are unknown and these are determined by applying the C^2-continuity condition.

For $n = 14$; $\theta = j/q$, $q = 50$ and $j = 0, \ldots, q$; $h_i = h1 = 1$; $r_i = r$, $t_i = t$ such that $t > r$; $d_1 = d_n = 0$ and for the data values given as:

$$\begin{aligned}\{(x_i, f_i)\}_{i=1}^{14} = \{&(122, 128); (122, 156); (150, 184); (178, 184);\\ &(206, 156); (206, 128); (178, 100); (150, 100);\\ &(122, 72); (122, 44); (150, 16); (178, 16);\\ &(206, 44); (206, 72)\},\end{aligned}$$

we obtain the C^2-rational cubic spline interpolant. Thus for different values of the tension parameters r and t, corresponding different graphs of the spline interpolant are obtained. From figure 1, we observe that as the values of the tension parameters r and t both are increased, we get considerable smooth curves of the C^2-rational (cubic/linear) spline interpolant. In fact, we observe the following:

(i) If only one parameter is taken into consideration as discussed in [6], i.e., if $r = 1$ and we have only one effective parameter t, then from figure 2 it can be seen that the curves obtained for different values of t are not smooth and as the value of t increases we get still worse curves.

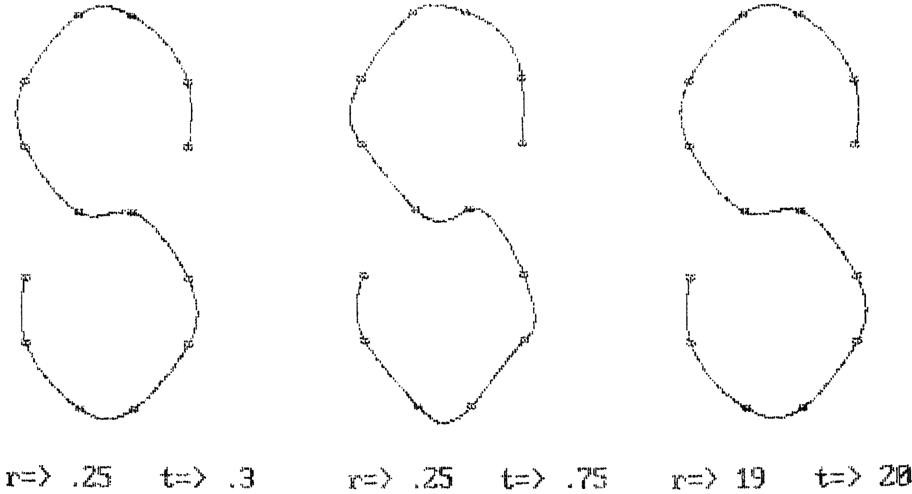


Figure 1.

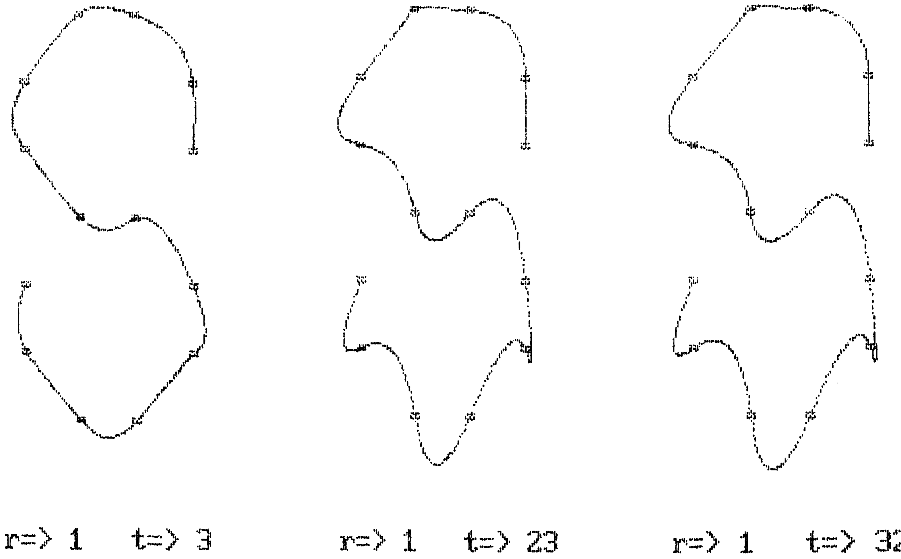


Figure 2.

(ii) From figures 3 and 4, we observe that when r and t are relatively different then
curves are not smooth.

(iii) When r and t are nearly equal and relatively larger in values then we have su
ciently smooth curves (see figures 1 and 4).

6. Remarks

6.1. The parameters r_i and t_i cannot be both zero since it leads to a trivial situation.
the choice $r_i = t_i$, the rational (cubic/linear) spline interpolant clearly reduces to us
cubic spline interpolating function values at the knots. This spline interpolant has b

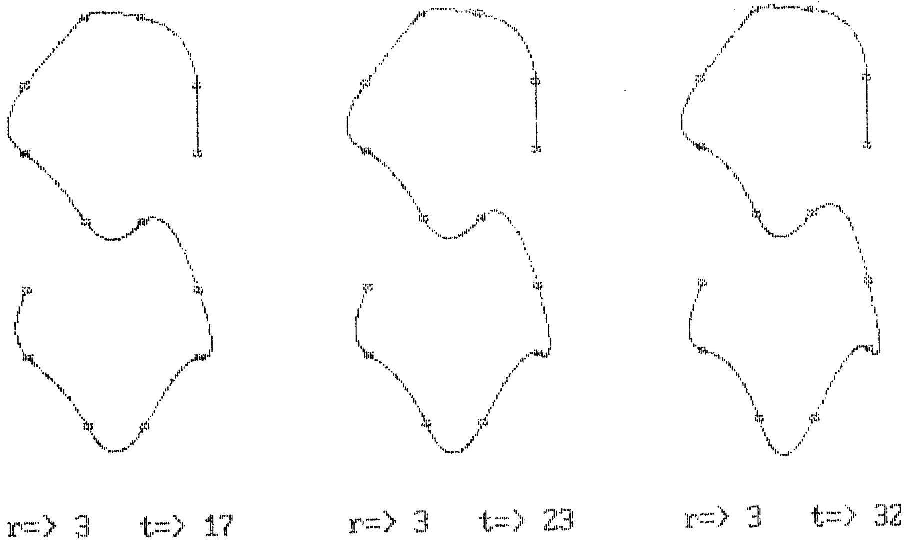


Figure 3.

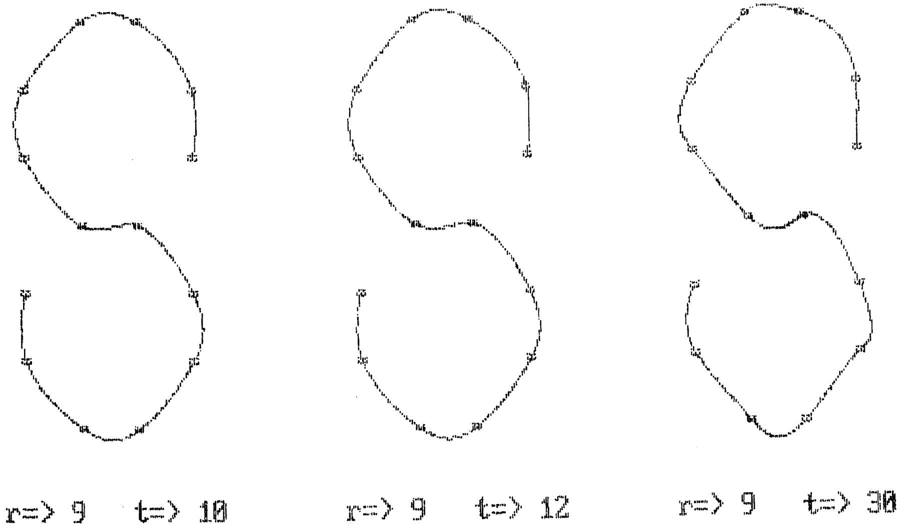


Figure 4.

studied in [4]. When either of the two parameters is zero, then the rational (cubic/linear) spline interpolant reduces to a rational (cubic/linear) interpolant which has been discussed in [7].

6.2. If the derivative parameters satisfy the inequality (3.6), then the C^1-rational (cubic/ linear) spline interpolant is monotonic as well as convex, provided (3.9) holds.

6.3. We observe that as $r_i \to t_i$ and $h_i \to 0$, then (2.16) reduces to

$$\|e\| \leq \frac{\bar{t}}{4t} h_i \max\{|d_i|, d_{i+1}|\}.$$

References

[1] Delbourgo R and Gregory J A, The determination of derivative parameters for a monotoni rational quadratic interpolant. *IMA J. Numer. Anal.* **5** (1985) 397–406

[2] Delbourgo R and Gregory J A, Shape preserving piecewise rational interpolation. *SIAM J. Sc Stat. Comput.* **6** (1985) 967–975

[3] Delbourgo R, Shape preserving interpolation to convex data by rational functions wit quadratic numerator and linear denominator. *IMA J. Numer. Anal.* **9** (1989) 123–136

[4] Fritsch F N and Carlson R E, Monotone piecewise cubic interpolation. *SIAM J. Numer. Ana* **17** (1980) 238–246

[5] Gregory J A and Delbourgo R, Piecewise rational quadratic interpolation to monotone dat *IMA J. Numer. Anal.* **2** (1982) 123–130

[6] Ismail M K, Monotonicity preserving interpolation using C^2-rational cubic Bezier curve *Mathematical Methods in CAGD II* (eds) T Lyche and L L Schumaker (1992) pp. 343–35

[7] Shrivastava M, Piecewise rational interpolation to monotonic and convex data. *J. Comp. App Math.* (communicated)

Proc. Indian Acad. Sci. (Math. Sci.), Vol. 110, No. 3, August 2000, pp. 315–322.
© Printed in India

Coin tossing and Laplace inversion

J C GUPTA

Indian Statistical Institute, New Delhi 110016, India
Current address: 32, Mirdha Tola, Budaun 243 601, India

MS received 5 May 1999; revised 3 April 2000

Abstract. An analysis of exchangeable sequences of coin tossings leads to inversion formulae for Laplace transforms of probability measures.

Keywords. Laplace inversion; moment problem; exchangeable probabilities.

1. Introduction

There is an intimate relationship between the Laplace transform

$$\phi(\lambda) = \int_0^\infty e^{-\lambda t} d\nu(t), \quad \lambda \geq 0 \tag{1.1}$$

of a probability measure ν on $[0, \infty)$ and the moment sequence

$$c(n) = \int_0^1 x^n d\mu(x), \quad n = 0, 1, 2, \ldots \tag{1.2}$$

of a probability measure μ on $(0, 1]$ via the obvious change of variables $e^{-t} = x$. An inversion formula for μ in terms of its moments yields an inversion formula for ν in terms of the values of its Laplace transform at $n = 0, 1, 2, \ldots$ and vice versa. In our discussion we allow μ (respectively ν) to have positive mass at 0 (respectively ∞).

Let $X_1, X_2, \ldots$ be 0, 1-valued random variables; one can identify 1 with 'heads' and 0 with 'tails'. These variables are said to be exchangeable if their joint distribution is invariant under finite permutations. Such variables can be generated in the following manner: first choose p at random according to a probability law μ on $[0, 1]$ and then let $X_1, X_2, \ldots$ be results of i.i.d tosses of a coin having probability p for 'heads'. The resulting measure

$$P(\cdot) = \int_0^1 P_p(\cdot) d\mu(p) \tag{1.3}$$

on $\{0, 1\}^N$ is a mixture of i.i.d probabilities P_p.

Then under P the process of coordinate functions is exchangeable. By a theorem of De Finetti, any exchangeable sequence of 0, 1-valued random variables arises in this manner for a suitable μ. The strong law of large numbers takes the form

$$Y_n := \frac{X_1 + X_2 + \cdots + X_n}{n} \longrightarrow Y_\infty \quad \text{a.s. } [P]. \tag{1.4}$$

Here the limit Y_∞ is a random variable. Further,

$$Y_\infty \sim \mu \quad \text{and} \quad \mathcal{L}(Y_n) \Longrightarrow \mu, \tag{1.5}$$

where $\mathcal{L}(Y_n)$ stands for the probability law of $Y_n, n \geq 1$ and '$\Longrightarrow$' denotes weak convergence. If $\{\tau_n\}_{n\geq 1}$ is a sequence of stopping times such that

$$\tau_n \rightarrow \infty \text{ a.s } [P], \qquad (1.6$$

then

$$Y_{\tau_n} \rightarrow Y_\infty \text{ a.s. } [P] \quad \text{and} \quad F_n \Longrightarrow F, \qquad (1.7$$

where $F, F_1, F_2, \ldots$ are the p.d.f.'s of $\mu, Y_{\tau_1}, Y_{\tau_2}, \ldots$ respectively. In the coin tossin situation many choices of $\{\tau_n\}_{n\geq 1}$ exist for which (1.6) holds and F_n can be explicitl written down in terms of $c(k)$, $k = 1, 2, \ldots$. Thus we get a host of inversion formulae fo μ in terms of its moments.

The classical inversion formulae, e.g., those due to Hausdorff, Widder, and Feller ca be obtained in the above manner. The methods of these authors were analytical althoug Feller was motivated by problems arising in stochastic theory of telephone traffic as h mentions in the introduction of his paper [2]. It is therefore satisfying to see that some c the results of Widder and Feller are consequences of the strong law of large numbers conditioning and stopping times, ideas which are central to probability theory.

This paper is organized as follows. Section 1 deals with the 1-dimensional case whil § 2 briefly deals with the 2-dimensional case; the generalization to higher dimensions i straightforward.

2. Coin tossing and inversion formulae

Exchangeable probabilities on the coin-tossing space

Let $E_1 = \{0, 1\}$ and $\Omega = E_1^N$ where $N = \{1, 2, 3, \ldots\}$; the space Ω is sometimes calle the coin-tossing space. Let $\omega = (\omega_1, \omega_2, \ldots)$ be a generic point of Ω and $X_1, X_2, \ldots$ be th coordinate variables, i.e., $X_n(\omega) := \omega_n$, $n = 1, 2, \ldots$. Let $\mathcal{F}_n = \sigma\langle X_1, X_2, \ldots, X_n\rangle$ an $\mathcal{F} = \sigma\langle X_n : n \geq 1\rangle$ be the σ-fields of subsets of Ω. Let Σ be the group of all permutatio of natural numbers which shift only finitely many of them and let $\Sigma_n = \{\sigma \in \Sigma : \sigma(i) =$ for $i > n\}$. For $\sigma \in \Sigma$, let $T_\sigma : \Omega \rightarrow \Omega$ be defined by $T_\sigma(\omega_1, \omega_2, \ldots) := (\omega_{\sigma(1)}, \omega_{\sigma(2)}, \ldots$ Let $\mathcal{S}_n = \{A \in \mathcal{F} : T_\sigma^{-1}A = A \text{ for all } \sigma \in \Sigma_n\}$ and $\mathcal{S} = \{A \in \mathcal{F} : T_\sigma^{-1}A = A$ for all σ $\Sigma\}$. Clearly $\mathcal{S}_n \downarrow \mathcal{S}$. A probability P on $(\Omega, \mathcal{F})$ is said to be exchangeable if PT_σ^{-1} P for all $\sigma \in \Sigma$. For each p, $0 \leq p \leq 1$, let P_p be the product probability on Ω und which $P_p(X_i = 1) = 1 - P_p(X_i = 0) = p$, $i = 1, 2, \ldots$. By a theorem of De Finetti, se e.g., Meyer [7], under an exchangeable P on $(\Omega, \mathcal{F})$, $X_1, X_2, \ldots$ are conditional independent given the symmetric σ-field $\mathcal{S}$. As a consequence, corresponding to exchangeable P there exists a probability μ on $[0, 1]$ such that

$$P(F) = \int_0^1 P_p(F)\mathrm{d}\mu(p), \quad F \in \mathcal{F}. \qquad (2.$$

The probability measure μ is called the *mixing probability* corresponding to P.

Inversion formulae

We fix a probability measure μ on $[0, 1]$ and the associated exchangeable probability P $(\Omega, \mathcal{F})$. Let

$$\Delta c(k) = c(k + 1) - c(k), \qquad (2.$$

where $c(k)$, $k = 0, 1, 2, \dots$ are given by (1.2). Then

$$(-1)^{n-k}\Delta^{n-k}c(k) = \int_0^1 (1-x)^{n-k}x^k \mathrm{d}\mu(x),\ k = 0, 1, 2, \dots, n \geq k. \tag{2.3}$$

The atoms of $\mathcal{S}_n$ are

$$\Delta_{n,k} = \{\omega \in \Omega; \#(i : 1 \leq i \leq n, \omega_i = 1) = k\} \tag{2.4}$$

and

$$P(\Delta_{n,k}) = (-1)^{n-k}\binom{n}{k}\Delta^{n-k}c(k). \tag{2.5}$$

It is easily seen that the mixing probability μ is uniquely determined by $P(\Delta_{n,k})$, $n = 1, 2, \dots,$ $k = 1, 2, \dots, n$.

As $\mathcal{S}_n \downarrow \mathcal{S}$, by the reverse martingale convergence theorem, we have

$$Y_n := E(X_1 \| \mathcal{S}_n) = \frac{X_1 + X_2 + \cdots + X_n}{n} \longrightarrow Y_\infty := E(X_1 \| \mathcal{S}) \text{ a.s. } [P]. \tag{2.6}$$

Further, for $k = 1, 2, \dots,$

$$\begin{aligned} Y_\infty^k &= E(X_1 \| \mathcal{S}) \cdot E(X_2 \| \mathcal{S}) \dots E(X_k \| \mathcal{S}) \text{ a.s. } [P] \\ &= E(X_1 X_2 \dots X_k \| \mathcal{S}) \text{ a.s. } [P] \text{ by De Finetti's theorem} \\ &= P(X_1 = X_2 = \cdots = X_k = 1 \| \mathcal{S}) \text{ a.s. } [P] \end{aligned}$$

and consequently,

$$E(Y_\infty^k) = \int_0^1 p^k \mathrm{d}\mu(p) = c(k).$$

Thus

$$Y_\infty \sim \mu. \tag{2.7}$$

Now let $\{\tau_n\}_{n \geq 1}$ be a sequence of stopping times with respect to $\{\mathcal{F}_n\}_{n \geq 1}$ such that

$$\tau_n \to \infty \text{ a.s. } [P]. \tag{2.8}$$

Then

$$Y_{\tau_n} \to Y_\infty \text{ a.s. } [P]. \tag{2.9}$$

By the compactness of $[0, 1]$, it follows by Prokhorov's theorem, see, e.g., [1], that the sequence of probability laws of Y_{τ_n} converges in the weak $*$-topology to μ. Thus

$$F_n \Longrightarrow F, \tag{2.10}$$

where $F, F_1, F_2, \dots$ are the p.d.f.'s of $\mu, Y_{\tau_1}, Y_{\tau_2}, \dots$ respectively.

For each choice of $\{\tau_n\}_{n \geq 1}$ for which (2.8) holds, we get an inversion formula for μ. We give some examples.

Example 1. Take $\tau_n \equiv n$. Then (2.8) holds and

$$P\left(Y_n = \frac{k}{n}\right) = P(\Delta_{n,k}) = (-1)^{n-k}\binom{n}{k}\Delta^{n-k}c(k) \text{ by (2.5).}$$

Thus

$$F_n(t) = \sum_{k:k\leq[nt]} (-1)^{n-k}\binom{n}{k}\Delta^{n-k}c(k) \to F(t)$$

at the points of continuity of F. This is the inversion formula of Hausdorff [3].

Example 2. Let τ_n be the waiting time for the appearance of the nth tail, i.e.,

$$\tau_n(\omega) = \inf\{m : X_1(\omega) + X_2(\omega) + \cdots + X_m(\omega) = m - n\};$$

we adopt the convention that the infimum of the empty set is ∞. Here $\tau_n(\omega) \geq n$ for all ω and (2.8) holds. Further,

$$P_p(\tau_n = n+k) = \binom{n+k-1}{k}p^k(1-p)^n, \quad k = 0,1,2,\ldots, \ 0 < p < 1,$$
$$P_0(\tau_n = n) = 1 \quad \text{and} \quad P_1(\tau_n = \infty) = 1.$$

Also,

$$P_p\left(Y_{\tau_n} = \frac{k}{n+k}\right) = \binom{n+k-1}{k}p^k(1-p)^n, \quad k = 0,1,2,\ldots, \ 0 < p < 1,$$
$$P_0(Y_{\tau_n} = 0) = 1 \quad \text{and} \quad P_1(Y_{\tau_n} = 1) = P_1(Y_\infty = 1) = 1$$

by the strong law of large numbers. Thus, by (2.1), the distribution function F_n places mass $(-1)^n\binom{n+k-1}{k}\Delta^n c(k)$ at $\frac{k}{k+n}$, $k = 0,1,2,\ldots$ and the remaining mass $\mu(\{1\}) = c(\infty)$ at 1.

This is essentially the inversion formula derived by Widder – see Theorem 42 and the footnote on p. 193 of [8]. The nth approximant of Widder places mass $(-1)^{n+1}\binom{n+k}{k}$ $\Delta^{n+1}c(k)$ at $\frac{k}{k+n}$, $k = 0,1,2,\ldots$ and mass $c(\infty)$ at 1 which agrees with our F_{n+1} except that $\frac{k}{k+n}$ is replaced by $\frac{k}{k+n+1}$, which hardly matters. It may be observed that F_n contains infinitely many jumps, they cluster at 1 and their amount uses differences of a fixed order.

Example 3. Let σ_n be the waiting time for the appearance of the nth head, i.e.,

$$\sigma_n(\omega) = \inf\{m : X_1(\omega) + X_2(\omega) + \cdots + X_m(\omega) = n\},$$

the infimum of the empty set being ∞. Here $\sigma_n(\omega) \geq n$ for all ω and (2.8) holds. Further,

$$P_p(\sigma_n = n+k) = \binom{n+k-1}{k}p^n(1-p)^k, \quad k = 0,1,2,\ldots, 0 < p < 1,$$
$$P_1(\sigma_n = n) = 1 \quad \text{and} \quad P_0(\sigma_n = \infty) = 1.$$

Also,

$$P_p\left(Y_{\sigma_n} = \frac{n}{n+k}\right) = \binom{n+k-1}{k}p^n(1-p)^k, \quad k = 0,1,2,\ldots, 0 < p < 1,$$
$$P_1(Y_{\sigma_n} = 1) = 1 \quad \text{and} \quad P_0(Y_{\sigma_n} = 0) = 1.$$

Thus, by (2.1), F_n places mass $(-1)^k\binom{n+k-1}{k}\Delta^k c(n)$ at $\frac{n}{n+k}$, $k = 0,1,2,\ldots$ and the remaining mass $\mu(\{0\}) = 1 - \Sigma_{k=0}^{\infty}(-1)^k\binom{n+k-1}{k}\Delta^k c(n)$ at 0. It may be observed that F_n contains infinitely many jumps, they cluster at 0 and their amount uses differences belonging to a fixed point.

The above formula was derived by very different methods by Feller – see theorem 5 and remark on pages 673–74 of [2]; in his case μ is supported on $(0,1]$ so that $\mu(\{0\})=0$.

Example 4. Let $\rho_n = \sigma_n \wedge \tau_n$ where τ_n and σ_n are as in examples 2 and 3 respectively. Then $\rho_n(\omega) \geq n$ for all ω and (2.8) holds. Further, $P_p(\rho_n = n+k) = \binom{n+k-1}{k}\{p^n q^k + p^k q^n\}$, $k = 0,1,2,\ldots,n-1$ and it is easily seen that F_n places mass $(-1)^k \binom{n+k-1}{k}\Delta^k c(n)$ at $\frac{n}{n+k}$ and mass $(-1)^n \binom{n+k-1}{k}\Delta^n c(k)$ at $\frac{k}{n+k}$, $k=0,1,2,\ldots,n-1$.

The reader is invited to choose his favourite $\{\tau_n\}_{n\geq 1}$ satisfying (2.8) and write down the corresponding sequence of approximants of the d.f. F of μ.

3. Inversion formulae in higher dimensions

We restrict ourselves to two dimensions; the generalization to higher dimensions is straightforward. The problem of inversion of the Laplace transform

$$\phi(\lambda_1,\lambda_2) = \int_0^\infty \int_0^\infty \mathrm{e}^{-(\lambda_1 t_1 + \lambda_2 t_2)} \mathrm{d}\nu(t_1,t_2), \quad \lambda_i \geq 0, i = 1,2 \tag{3.1}$$

of a probability measure ν on $[0,\infty)\times[0,\infty)$ in terms of $\phi(k,\ell), k,\ell = 0,1,2,\ldots$ is same as that of finding an inversion formula for a probability measure μ on $(0,1]\times(0,1]$ in terms of its moments

$$c(k,\ell) = \int_0^1 \int_0^1 x_1^k x_2^\ell \mathrm{d}\mu(x_1,x_2). \tag{3.2}$$

In our discussion we consider probability measures μ on $I^2 := [0,1]\times[0,1]$.

To do an analysis similar to the 1-dimensional case, we introduce a special kind of exchangeable probability on the space of a sequence of tosses of a pair of coins. First we set up the notation. Let $E_2 = \{(00),(01),(10),(11)\}$, $\Omega = E_2^{\mathbb{N}}$, $\omega = (\omega_1,\omega_2,\ldots)$ with $\omega_i = (\omega_{i1}\,\omega_{i2})$, $i=1,2,\ldots$, be a generic point of Ω and $X_n = (X_{n1},X_{n2})$ with $X_{n1}(\omega) = \omega_{n1}$, $X_{n2}(\omega) = \omega_{n2}, n = 1,2,\ldots$, be the coordinate variables. Let $\mathcal{F}_n, \mathcal{F}, \Sigma$, Σ_n, $T_\sigma, \mathcal{S}_n$ and $\mathcal{S}$ be defined as before. For $p=(p_1,p_2)\in I^2$ let θ_p be the probability on E_2 defined by

$$\begin{aligned}&\theta_p(00) = (1-p_1)(1-p_2),\ \theta_p(01) = (1-p_1)p_2,\\ &\theta_p(10) = p_1(1-p_2) \text{ and } \theta_p(11) = p_1p_2\end{aligned} \tag{3.3}$$

and let $P_p = \theta_p \times \theta_p \times \ldots$ be the corresponding product probability on Ω. We fix a probability μ on I^2 and introduce the exchangeable probability P on $(\Omega,\mathcal{F})$ by

$$P(F) = \int_{I^2} P_p(F)\mathrm{d}\mu(p), \quad F\in\mathcal{F}. \tag{3.4}$$

Let

$$\begin{aligned}&\Delta_1 c(k,\ell) := c(k+1,\ell) - c(k,\ell) \quad \text{and}\\ &\Delta_2 c(k,\ell) := c(k,\ell+1) - c(k,\ell),\end{aligned} \tag{3.5}$$

where $c(k,\ell), k, \ell = 0,1,2,\ldots$ are given by (3.2). We have

$$(-1)^{2n-k-\ell}\Delta_1^{n-k}\Delta_2^{n-\ell}c(k,\ell) = \int_{I^2} x_1^k x_2^\ell (1-x_1)^{n-k}(1-x_2)^{n-\ell}\mathrm{d}\mu(x_1,x_2). \tag{3.6}$$

The probability of the atoms of $\mathcal{S}_n$ can be written in terms of these differences and it is easily seen that the mixing probability μ is uniquely determined by the values of P on the atoms of $\mathcal{S}_n, n \geq 1$.

By De Finetti's theorem, under $P, X_n, n \geq 1$ are conditionally independent given $\mathcal{S}$; further, by our construction, the mixing probability μ is supported on the set of probabilities of type θ_p on E_2 as given in (3.3). Therefore

(a) for almost all $\omega[P]$, $(E(X_{n1}\|\mathcal{S})(\omega), E(X_{n2}\|\mathcal{S})(\omega)), n = 1,2,\ldots$ are like tosses of a pair of coins having probability of 'heads', say $p_1(\omega)$ and $p_2(\omega)$, *all* the tosses being independent and

(b) $(p_1,p_2) \sim \mu$.

By the reverse martingale convergence theorem,

$$\begin{aligned} Y_{n1} &:= E(X_{11}\|\mathcal{S}_n) = \frac{X_{11}+X_{21}+\cdots+X_{n1}}{n} \to Y_{\infty 1} := E(X_{11}\|\mathcal{S}) \text{ a.s. } [P] \\ Y_{n2} &:= E(X_{12}\|\mathcal{S}_n) = \frac{X_{12}+X_{22}+\cdots+X_{n2}}{n} \to Y_{\infty 2} := E(X_{12}\|\mathcal{S}) \text{ a.s. } [P]. \end{aligned} \tag{3.7}$$

Further, by (a) and (b) above, for $k,\ell = 0,1,2,\ldots$, we have

$$\begin{aligned} Y_{\infty 1}^k Y_{\infty 2}^\ell &= \{\Pi_{i=1}^k E(X_{i1}\|\mathcal{S})\}\{\Pi_{j=1}^\ell E(X_{j2}\|\mathcal{S})\} \text{ a.s. } [P] \\ &= E\{(\Pi_{i=1}^k X_{i1})(\Pi_{j=1}^\ell X_{j2})\|\mathcal{S}\} \text{ a.s. } [P] \\ &= P(X_{11} = X_{21} = \cdots = X_{k1} = X_{12} = X_{22} = \cdots = X_{\ell 2} = 1\|\mathcal{S}) \text{ a.s. } [P] \\ &= p_1^k p_2^\ell \end{aligned}$$

and

$$\begin{aligned} E(Y_{\infty 1}^k Y_{\infty 2}^\ell) &= \int_{I^2} p_1^k p_2^\ell \mathrm{d}\mu(p) \\ &= c(k,\ell). \end{aligned}$$

Thus

$$(Y_{\infty 1}, Y_{\infty 2}) \sim \mu. \tag{3.8}$$

Now let $\{\tau_n\}_{n\geq 1}$ and $\{\sigma_n\}_{n\geq 1}$ be two sequences of stopping times with respect to $\{\mathcal{F}_n\}_{n\geq 1}$ such that

$$\tau_n \to \infty \text{ a.s. } [P], \quad \sigma_n \to \infty \text{ a.s. } [P]. \tag{3.9}$$

Then, by (3.7) and (3.9),

$$(Y_{\tau_n 1}, Y_{\sigma_n 2}) \to (Y_{\infty 1}, Y_{\infty 2}) \text{ a.s. } [P]. \tag{3.10}$$

3.8), (3.10) and the compactness of I^2 it follows that the sequence of probability laws $Y_{\tau_n1}, Y_{\sigma_n2})$ converges in the weak $*$-topology to μ. Thus

$$G_n \Longrightarrow G, \tag{3.11}$$

re $G, G_1, G_2, \dots$ are the p.d.f.'s of μ, $(Y_{\tau_11}, Y_{\sigma_12})$, $(Y_{\tau_21}, Y_{\sigma_22}), \dots$ respectively.
or each choice of $\{\tau_n\}_{n\geq 1}$ and $\{\sigma_n\}_{n\geq 1}$ for which (3.9) holds we get an inversion nula for μ. We give some examples.

mple 1. Let $\tau_n \equiv n$ and $\sigma_n \equiv n$. Then (3.9) holds and

$$P\left(Y_{n1} = \frac{k}{n}, Y_{n2} = \frac{\ell}{n}\right) = (-1)^{2n-k-\ell}\Delta_1^{n-k}\Delta_2^{n-\ell}c(k,\ell) \text{ by (3.6)},$$

$= 0, 1, 2, \dots, n$. Thus

$$G_n(s,t) = \sum_{\substack{k:k\leq[ns]\\ \ell:\ell\leq[nt]}} (-1)^{2n-k-\ell}\Delta_1^{n-k}\Delta_2^{n-\ell}c(k,\ell) \to G(s,t)$$

ne points of continuity of G. This inversion formula can be found in [6]; also see [4] [5].

mple 2. Let τ_n (respectively σ_n) be the waiting time for the appearance of the nth tail the first (respectively second) coin, i.e.,

$$\tau_n(\omega) = \inf\{m : X_{11}(\omega) + X_{21}(\omega) + \cdots + X_{m1}(\omega) = m - n\},$$
$$\sigma_n(\omega) = \inf\{m : X_{12}(\omega) + X_{22}(\omega) + \ldots + X_{m2}(\omega) = m - n\},$$

infimum of an empty set being ∞. Then (3.9) holds and for $k, \ell = 0, 1, 2, \dots,$

$$\begin{aligned} P_{(p_1,p_2)}\left\{Y_{\tau_n1} = \frac{k}{n+k}, Y_{\sigma_n2} = \frac{\ell}{n+\ell}\right\} &= P_{(p_1,p_2)}\{\tau_n = n+k, \sigma_n = n+\ell\} \\ &= \binom{n+k-1}{k}\binom{n+\ell-1}{\ell}p_1^k p_2^\ell(1-p_1)^n(1-p_2)^n, 0 \leq p_1, p_2 < 1, \\ P_{(1,p_2)}\left\{Y_{\tau_n1} = 1, Y_{\sigma_n2} = \frac{\ell}{n+\ell}\right\} &= P_{(1,p_2)}\{\tau_n = \infty, \sigma_n = n+\ell\} \\ &= \binom{n+\ell-1}{\ell}p_2^\ell(1-p_2)^n, \quad 0 \leq p_2 < 1, \\ P_{(p_1,1)}\left\{Y_{\tau_n1} = \frac{k}{n+k}, Y_{\sigma_n2} = 1\right\} &= P_{(p_1,1)}\{\tau_n = n+k, \sigma_n = \infty\} \\ &= \binom{n+k-1}{k}p_1^k(1-p_1)^n, \quad 0 \leq p_1 < 1 \end{aligned}$$

$$P_{(1,1)}\{Y_{\tau_n1} = 1, Y_{\sigma_n2} = 1\} = P_{(1,1)}\{\tau_n = \infty, \sigma_n = \infty\} = 1.$$

p.d.f. G_n of $(Y_{\tau_n1}, Y_{\sigma_n2})$ places mass $(-1)^{2n}\binom{n+k-1}{k}\binom{n+\ell-1}{\ell}\Delta_1^n\Delta_2^n c(k,\ell)$ at $\left(\frac{k}{n+k}, \frac{\ell}{n+\ell}\right)$, $)^n\binom{n+\ell-1}{\ell}\Delta_2^n c(\infty,\ell)$ at $\left(1, \frac{\ell}{n+\ell}\right)$, $(-1)^n\binom{n+k-1}{k}\Delta^n c(k,\infty)$ at $\left(\frac{k}{n+k}, 1\right)$ and $c(\infty,\infty)$ at

(1,1). Here $c(\infty,\ell)$ stands for $\lim_{m\to\infty} c(m,\ell)$, $c(k,\infty)$ for $\lim_{m\to\infty} c(k,m)$ and $c(\infty,\infty) = \lim_{m\to\infty} c(m,m)$. This gives an inversion formula which is a 2-dimensional analogue of Widder's formula.

Example 3. Let τ_n (respectively σ_n) be the waiting time for the appearance of nth head or tail, whichever is earlier, for the first (respectively second) coin. Then (3.9) holds and it is easily seen that G_n places mass

$$(-1)^{k+\ell}\binom{n+k-1}{k}\binom{n+\ell-1}{\ell}\Delta_1^k\Delta_2^\ell c(n,n) \text{ at } \left(\frac{n}{n+k},\frac{n}{n+\ell}\right),$$
$$(-1)^{k+n}\binom{n+k-1}{k}\binom{n+\ell-1}{\ell}\Delta_1^k\Delta_2^n c(n,\ell) \text{ at } \left(\frac{n}{n+k},\frac{\ell}{n+\ell}\right),$$
$$(-1)^{n+\ell}\binom{n+k-1}{k}\binom{n+\ell-1}{\ell}\Delta_1^n\Delta_2^\ell c(k,n) \text{ at } \left(\frac{k}{n+k},\frac{n}{n+\ell}\right)$$

and

$$(-1)^{2n}\binom{n+k-1}{k}\binom{n+\ell-1}{\ell}\Delta_1^n\Delta_2^n c(k,\ell) \text{ at } \left(\frac{k}{n+k},\frac{\ell}{n+\ell}\right),$$
$$k,\ell = 0,1,2,\ldots,n-1.$$

The reader is invited to choose his favourite $\{\tau_n\}_{n\geq 1}$ and $\{\sigma_n\}_{n\geq 1}$ for which (3.9) holds and write down the corresponding inversion formula.

Acknowledgements

The author thanks the Indian Statistical Institute, New Delhi Centre, for the facilities extended to him. The author also thanks an anonymous referee for his suggestions.

References

[1] Billingsley P, *Convergence of Probability Measures,* (New York: John Wiley) (1968)

[2] Feller Willy, Completely monotone functions and sequences. *Duke Math. J.* **5** (1939) 661–674

[3] Hausdorff F, Momentprobleme für ein endliches Intervall, *Math. Z.* **16** (1923) 220–248

[4] Haviland E K, On the momentum problem for distribution functions in more than one dimension, *Am. J. Math.* **57** (1935) 562–568

[5] Haviland E K, On the momentum problem for distribution functions in more than one dimension II, *Am. J. Math.* **58** (1936) 164–168

[6] Hildebrandt T H and Schoenberg I J, On linear functional operations and the moment problem for a finite interval in one or several dimensions, *Ann. Math.* **34** (1933) 317–328

[7] Meyer Paul A, *Probability and Potentials* (Boston: Blaisdell Publishing Co.) (1966)

[8] Widder D V, The inversion of the Laplace integral and the related moment problem, *Trans. Am. Math. Soc.* **36** (1934) 107–200

Proc. Indian Acad. Sci. (Math. Sci.), Vol. 110, No. 3, August 2000, pp. 323–334.

Differential equations related to the Williams–Bjerknes tumour model

F MARTINEZ and A R VILLENA*

Departamento de Estaditica e I.O.; *Departamento de Analisis Matematico, Facultad de Ciencias, Universidad de Granada. 18071 Granada, Spain
E-mail: falvarez@goliat.ugr.es; avillena@goliat.ugr.es

MS received 26 October 1999

Abstract. We investigate an initial value problem which is closely related to the Williams–Bjerknes tumour model for a cancer which spreads through an epithelial basal layer modeled on $I \subset Z^2$. The solution of this problem is a family $p = (p_i(t))$, where each $p_i(t)$ could be considered as an approximation to the probability that the cell situated at i is cancerous at time t. We prove that this problem has a unique solution, it is defined on $[0, +\infty[$, and, for some relevant situations, $\lim_{t\to\infty} p_i(t) = 1$ for all $i \in I$. Moreover, we study the expected number of cancerous cells at time t.

Keywords. Williams–Bjerknes tumour model; expected number of cancerous cells.

1. Introduction

Based on chemical tests and mitotic patterns, Williams and Bjerknes proposed in [8] a model for the cancer spread through an epithelial basal layer. Independently, the Williams–Bjerknes tumour model was formulated within the field of interacting particle systems as the biased voter model (see [7]). Assuming that cells are situated on the lattice Z^2, the set ξ_t^A of sites occupied by cancerous cells at time t, given that the initial state is A, is a Markov process on the state space $\{\text{finite subsets of } Z^2\}$ thoroughly studied in [1–3,6].

In [5] we studied the spread of cancerous cells through the basal layer of an epithelium modeled on the lattice $I = Z^2$. In that paper we derived differential inequalities for the family $(p_i)_{i\in I}$, where we write $p_i(t)$ for the probability that the cell situated at i is cancerous at time t.

The aim of this paper is to study the following problem:

$$p_i' = -p_i + \frac{\kappa}{\omega_i}\sum_{\|j-i\|_1=1} p_j - (\kappa-1)p_i\frac{1}{\omega_i}\sum_{\|j-i\|_1=1} p_j, \quad \forall i \in I, \tag{1.1}$$

$$p_i(0) = a_i \quad \forall i \in I, \tag{1.2}$$

$$0 \le p_i \le 1 \quad \forall i \in I, \tag{1.3}$$

where I is assumed to be an arbitrary (finite or infinite) nonempty subset of Z^2, for each $i \in I$ the neighbours of the cell located at i are the elements of the set

$$\Omega_i = \{j \in I\colon \|i-j\|_1 = |i_1 - j_1| + |i_2 - j_2| = 1\}$$

whose cardinality is $\omega_i(\le 4)$, and $(a_i)_{i\in I}$ is the initial state of the epithelium. We assume that for $i,j \in I$ there is a sequence of sites $i_0 = i, i_1, \ldots, i_{n-1},\ i_n = j$ with $i_k \in \Omega_{k-1}$ for

$k = 1, \ldots, n$. A computer simulation shows that the solution of the preceding problem provides a good approximation to the Williams–Bjerknes tumour model.

We prove that this problem has a unique solution, it is defined on $[0, +\infty[$, and, for some specially relevant situations, $\lim_{t\to+\infty} p_i(t) = 1\ \forall i \in I$.

2. The tumour growth model

Cells are assumed to be of two types, normal and cancerous, and are located on a suitable lattice, one at each site. With each cellular division, one daughter remains in the site, while the other displaces a neighbouring cell which is pushed out of the basal layer. Cancerous cells are assumed to divide at a faster rate than normal cells. Splitting times for both normal and cancerous cells are assumed to be independent and have exponential distributions with parameter 1 and $\kappa > 1$, respectively. This makes the probability that a normal cell will split in the time interval $[t, t + \Delta t]$ equals Δt, irrespective of the time since its last division. For the cancerous cells, this event occurs with probability $\kappa\Delta t$.

It is easily seen that the probability of the cell i to be cancerous at time $t + \Delta t$ can be expressed as

$$\begin{aligned} p_i(t+\Delta t) =& p_i(t)\left[1 - \frac{\Delta t}{\omega_i}\sum_{j\in\Omega_i}(1 - u_j(t))\right] + o(\Delta t) \\ &+ (1 - p_i(t))\left[\frac{\kappa\Delta t}{\omega_i}\sum_{j\in\Omega_i} v_j(t)\right] + o(\Delta t), \end{aligned}$$

where we write $u_j(t)$ for the conditional probability that the cell located at j is cancerous at time t given that the cell located at i is cancerous and $v_j(t)$ stands for the conditional probability that the cell located at j is cancerous at time t given that the cell located at i is normal. Consequently, we have

$$\frac{p_i(t+\Delta t) - p_i(t)}{\Delta t} = -p_i(t)\frac{1}{\omega_i}\sum_{j\in\Omega_i}(1 - u_j(t)) + (1 - p_i(t))\frac{\kappa}{\omega_i}\sum_{j\in\Omega_i} v_j(t) + \frac{o(\Delta t)}{\Delta t}.$$

Assume that the functions p_i are differentiable on $[0, \tau[$ for some $\tau > 0$ for all $i \in I$ and let Δt approach zero. This yields

$$p_i'(t) = -p_i(t)\frac{1}{\omega_i}\sum_{j\in\Omega_i}(1 - u_j(t)) + (1 - p_i(t))\frac{\kappa}{\omega_i}\sum_{j\in\Omega_i} v_j(t).$$

Since $p_j(t) = u_j(t)p_i(t) + v_j(t)(1 - p_i(t))$, we get

$$\sum_{j\in\Omega_i} p_j(t) = \sum_{j\in\Omega_i} u_j(t)p_i(t) + \sum_{j\in\Omega_i} v_j(t)(1 - p_i(t))$$

and therefore

$$p_i'(t) = -p_i(t)\frac{1}{\omega_i}\sum_{j\in\Omega_i}(1 - u_j(t)) + \frac{\kappa}{\omega_i}\left(\sum_{j\in\Omega_i} p_j(t) - \sum_{j\in\Omega_i} u_j(t)p_i(t)\right).$$

To get a closed system of equations one needs to obtain equations for the functions u_j. These equations involve other still higher order correlation functions. Here we replace each u_j by p_j.

3. Existence and uniqueness of solutions

Let $l_\infty(I)$ denote the set of all real families $x = (x_i)_{i\in I}$ for which $\|x\|_\infty = \sup\limits_{i\in I}|x_i| < +\infty$, which becomes a Banach space with pointwise operations and the supremum norm. For every $i \in I$ let E_i denote the continuous linear functional on $l_\infty(I)$ given by $E_i(x) = x_i$. We will denote by P the closed ball

$$\left\{x \in l_\infty(I)\colon \sup_{i\in I}|x_i - 1/2| \le 1/2\right\} = \{x \in l_\infty(I)\colon 0 \le x_i \le 1 \ \forall i \in I\}$$

whose interior is

$$\begin{aligned}\operatorname{int}(P) &= \left\{x \in l_\infty(I)\colon \sup_{i\in I}|x_i - 1/2| < 1/2\right\}\\ &= \bigcup_{0<\rho<\frac{1}{2}} \{x \in l_\infty(I)\colon \rho \le x_i \le 1-\rho, \forall i \in I\}.\end{aligned}$$

Let f be the function from $l_\infty(I)$ into itself defined by

$$E_i(f(x)) = -x_i + \frac{\kappa}{\omega_i}\sum_{j\in\Omega_i} x_j - (\kappa-1)x_i\frac{1}{\omega_i}\sum_{j\in\Omega_i} x_j$$

for all $x \in l_\infty(I)$ and $i \in I$. Let $a \in l_\infty(I)$, and consider the initial value problem on $l_\infty(I)$

$$\begin{cases} x' &= f(x)\\ x(0) &= a.\end{cases} \tag{2}$$

It is clear that if a family of functions $(p_i)_{i\in I}$ is a solution of (2) with $a \in P$, then it satisfies (1.1) and (1.2), and it is well-known that the converse is true for a finite I. The problem is that, for an infinite I, a family of differentiable functions $(p_i)_{i\in I}$ on the interval $[0,\tau[$ may be not differentiable as a $l_\infty(I)$-valued function. This difficulty is solved by our next theorem.

Theorem 1. *Let $a \in P$ and let $(p_i)_{i\in I}$ be a family of functions from the interval $[0,\tau[$ into $[0,1]$. Then the following assertions are equivalent:*

(i) *$(p_i)_{i\in I}$ satisfies (1.1) and (1.2).*
(ii) *The function p from $[0,\tau[$ into $l_\infty(I)$ defined by $p(t) = (p_i(t))_{i\in I}$ for all $t \in [0,\tau[$ is a solution of (2) on $[0,\tau[$.*

Proof. If (ii) obtains, then the chain rule shows that the functions $p_i = E_i \circ p$ are differentiable on $[0,\tau[$ with $p_i'(t) = E_i(p'(t)) = E_i(f(p(t)))$ for all $i \in I$. Consequently, $(p_i)_{i\in I}$ satisfies (1.1). (1.2) is obvious.

Assume that (i) holds. Then it is clear that, for all $i \in I$, the function p_i is infinitely differentiable on $[0,\tau[$. Moreover we have

$$\begin{aligned}|p_i'(t)| &\le |p_i(t)| + \frac{\kappa}{\omega_i}\sum_{j\in\Omega_i}|p_j(t)| + (\kappa-1)|p_i(t)|\frac{1}{\omega_i}\sum_{j\in\Omega_i}|p_j(t)|\\ &\le 1+\kappa+(\kappa-1) = 2\kappa\end{aligned}$$

and

$$\begin{aligned}|p_i''(t)| &\leq |p_i'(t)| + \frac{\kappa}{\omega_i}\sum_{j\in\Omega_i}|p_j'(t)| + (\kappa-1)|p_i'(t)|\frac{1}{\omega_i}\sum_{j\in\Omega_i}|p_j(t)|\\ &\quad + (\kappa-1)|p_i(t)|\frac{1}{\omega_i}\sum_{j\in\Omega_i}|p_j'(t)|\\ &\leq 2\kappa + \kappa(2\kappa) + (\kappa-1)(2\kappa) + (\kappa-1)(2\kappa) = 6\kappa^2 - 2\kappa\end{aligned}$$

for all $t \in [0,\tau[$ and $i \in I$.

Let $i \in I$, $t \in [0,\tau[$ and $0 \neq h \in R$ such that $t+h \in [0,\tau[$. The mean value theorem provides $\xi, \zeta \in R^+$ such that $|\xi - t| \leq |h|$, $|\zeta - t| < |\xi - t|$, $p_i(t+h) - p_i(t) = p_i'(\xi)h$, and $p_i'(\xi) - p_i'(t) = p_i''(\zeta)(\zeta - t)$. Hence

$$\begin{aligned}\left|\frac{p_i(t+h)-p_i(t)}{h} - p_i'(t)\right| &= |p_i'(\xi) - p_i'(t)| = |p_i''(\zeta)(\xi - t)|\\ &\leq (6\kappa^2 - 2\kappa)|\xi - t| \leq (6\kappa^2 - 2\kappa)|h|\end{aligned}$$

and therefore

$$\left\|\frac{p(t+h)-p(t)}{h} - (p_i')_{i\in I}\right\|_\infty \leq (6\kappa^2 - 2\kappa)|h|.$$

This shows that p is differentiable on $[0,\tau[$ with $p'(t) = (p_i'(t))_{i\in I}$ for all $t \in [0,\tau[$ and consequently p is a solution of (2). □

In order to apply the classical existence theorem to the problem (2) we first show that the function f is infinitely differentiable.

Lemma 1. *The function f is infinitely differentiable on $l_\infty(I)$ and*

$$\|Df(x)\| \leq 1 + \kappa + 2(\kappa - 1)\|x\|_\infty$$

for all $x \in l_\infty(I)$.

Proof. Let $x \in l_\infty(I)$ and T_x be the linear operator from $l_\infty(I)$ into itself defined by

$$E_iT_x(u) = -u_i + \frac{\kappa}{\omega_i}\sum_{j\in\Omega_i}u_j - (\kappa-1)x_i\frac{1}{\omega_i}\sum_{j\in\Omega_i}u_j - (\kappa-1)u_i\frac{1}{\omega_i}\sum_{j\in\Omega_i}x_j$$

for all $u \in l_\infty(I)$ and $i \in I$. We see at once that

$$|E_iT_xu| \leq (1 + \kappa + 2(\kappa-1)\|x\|_\infty)\|u\|_\infty$$

for all $u \in l_\infty(I)$ and $i \in I$, and so

$$\|T_xu\|_\infty \leq (1 + \kappa + 2(\kappa-1)\|x\|_\infty)\|u\|_\infty$$

for all $u \in l_\infty(I)$. Therefore T_x is continuous and

$$\|T_x\| \leq 1 + \kappa + 2(\kappa-1)\|x\|_\infty.$$

Furthermore, for all $i \in I$, we have

$$E_i(f(x+u) - f(x) - T_x(u)) = -(\kappa-1)u_i\frac{1}{\omega_i}\sum_{j\in\Omega_i}u_j$$

nd consequently

$$\|f(x+u)-f(x)-T_x(u)\|_\infty \leq (\kappa-1)\|u\|_\infty^2$$

or all $u \in l_\infty(I)$. Hence f is differentiable at x with $Df(x) = T_x$.

It is clear that Df is differentiable on $l_\infty(I)$ and $D^2 f$ is easily seen to satisfy

$$E_i(D^2 f(x)(u,v)) = -(\kappa-1)u_i \frac{1}{\omega_i}\sum_{j\in\Omega_i} v_j - (\kappa-1)v_i\frac{1}{\omega_i}\sum_{j\in\Omega_i} u_j$$

or all $x, u, v \in l_\infty(I)$.

Since $D^2 f$ is constant, we conclude that $D^3 f = 0$. □

From the preceding result and ([4]; 10.4.5 and 10.5.2) it may be concluded the ollowing.

emma 2. *If $a \in l_\infty(I)$, then* (2) *has a unique noncontinuable solution which is defined n an interval* $]-\sigma,\tau[$ *with* $\sigma, \tau > 0$.

From now on, by a solution of (2) we mean a solution of (2) on the interval $[0,\tau[$ hich is noncontinuable to the right.

emma 3. *If $a \in \operatorname{int}(P)$, then the solution p of* (2) *is defined on* $[0,+\infty[$ *and there exists* $\in]0,1/2[$ *such that* $\rho e^{-t} \leq p_i(t) \leq 1-\rho e^{-\kappa t}$ *for all* $t \in [0,+\infty[$ *and* $i \in I$.

roof. Let $\rho \in]0,1/2[$ such that $\rho \leq a_i \leq 1-\rho$ for all $i \in I$.

Let $[0,\tau[$ be the interval of existence of p and set

$$\sigma = \sup\{s \in [0,\tau[: p(t) \in \operatorname{int}(P), \forall t \in [0,s[\}.$$

rom ([4]; 10.4.5 and 10.5.2) we see that the initial value problem (2) in $\operatorname{int}(P)$ has a nique noncontinuable solution whose restriction to $[0,+\infty[$ is obviously the restriction f p to the interval $[0,\sigma[$.

Fix $i \in I$. Since $0 < p_i(t) < 1\ \forall t \in [0,\sigma[$, we see that

$$\frac{\mathrm{d}(\exp(t)p_i(t))}{\mathrm{d}t} = \exp(t)(1-p_i(t))\frac{\kappa}{\omega_i}\sum_{j\in\Omega_i} p_j(t) + \exp(t)p_i(t)\frac{1}{\omega_i}\sum_{j\in\Omega_i} p_j(t) \geq 0$$

or all $t \in [0,\sigma[$. Therefore the function $\exp(t)p_i(t)$ is strictly increasing on $[0,\sigma[$. From his we have

$$\rho \leq a_i = p_i(0) \leq \exp(t)p_i(t),$$

$t \in [0,\sigma[$. On the other hand, we have

$$p_i'(t) = \kappa(1-p_i(t))\frac{1}{\omega_i}\sum_{j\in\Omega_i} p_j(t) - p_i(t)\left[1-\frac{1}{\omega_i}\sum_{j\in\Omega_i} p_j(t)\right] \leq \kappa(1-p_i(t))$$

nd therefore $p_i'(t)/(1-p_i(t)) \leq \kappa\ \forall t \in [0,\sigma[$. Consequently,

$$\ln\left(\frac{1-p_i(0)}{1-p_i(t)}\right) = \int_0^t \frac{p_i'(s)}{1-p_i(s)}\mathrm{d}s \leq \int_0^t \kappa \mathrm{d}s = \kappa t,$$

which gives

$$p_i(t) \leq 1 - (1 - a_i)\exp(-\kappa t) \leq 1 - \rho \exp(-\kappa t)$$

for all $t \in [0, \sigma[$. Hence

$$\rho \exp(-\kappa\sigma) \leq \rho \exp(-t) \leq p_i(t) \leq 1 - \rho \exp(-\kappa t) \leq 1 - \rho \exp(-\kappa\sigma)$$

for all $t \in [0, \sigma[$ and $i \in I$.

Thus the closure of the set $p([0, \sigma[)$ in $l_\infty(I)$ is contained in the set

$$\{x \in l_\infty(I) \colon \rho \exp(-\kappa\sigma) \leq x_i \leq 1 - \rho \exp(-\kappa\sigma) \forall i \in I\} \subset \operatorname{int}(P).$$

Furthermore, it is immediate that $\|f(p(t))\|_\infty \leq 2\kappa$ for all $t \in [0, \sigma[$. According to ([4]; 10.5.5 and 10.5.5.1), we have $\sigma = +\infty$ and, in consequence, $\tau = +\infty$. □

Theorem 2. *If $a \in P$, then* (2) *has a unique solution p and this solution satisfies the following properties:*

(i) *p is defined on $[0, +\infty[$.*

(ii) *$0 \leq p_i(t) \leq 1$ for all $t \in [0, +\infty[$ and $i \in I$. Accordingly, p is the unique family of differentiable functions on $[0, +\infty[$ satisfying* (1.1), (1.2), *and* (1.3).

(iii) *If there are $i_0 \in I$ and $t_0 \in]0, +\infty[$ such that $p_{i_0}(t_0) = 0$, then $p_i(t) = 0$ for all $t \in [0, +\infty[$ and $i \in I$. If there are $i_0 \in I$ and $t_0 \in]0, +\infty[$ such that $p_{i_0}(t_0) = 1$, then $p_i(t) = 1$ for all $t \in [0, +\infty[$ and $i \in I$.*

Proof. Let $\{a_n\}$ be a sequence in $\operatorname{int}(P)$ such that $a = \lim a_n$. Let p_n be the unique solution of (2) with $p_n(0) = a_n$.

From Lemma 1 we see that $\|Df(x)\| \leq 3\kappa - 1 \ \forall x \in \operatorname{int}(P)$. On account of the preceding lemma and ([4]; 10.5.1), we have

$$\|p_m(t) - p_n(t)\|_\infty \leq \|a_m - a_n\|_\infty e^{(3\kappa-1)t}$$

for all $t \in [0, +\infty[$ and $m, n \in N$. It follows that $\{p_n\}$ is uniformly Cauchy on every compact subset of $[0, +\infty[$ and so $\{p_n\}$ converges uniformly on every compact subset of $[0, +\infty[$ to a function q. Since $\{p_n'\} = \{f(p_n)\}$ converges uniformly on every compact subset of $[0, +\infty[$ to $f(q)$, it may be concluded that q is differentiable on $[0, +\infty[$ with $q' = f(q)$. As $q(0) = \lim p_n(0) = \lim a_n = a$ we have $p = q$, which gives (i).

Lemma 3 leads to $p_n(t) \in P$ for all $t \in [0, +\infty[$ and $n \in N$ and therefore $p(t) = \lim p_n(t) \in P$ for all $t \in [0, +\infty[$. Theorem 1 now shows that p is the unique family of differentiable functions on $[0, +\infty[$ satisfying (1.1), (1.2), and (1.3).

Assume that there exist $i_0 \in I$ and $t_0 \in]0, +\infty[$ such that $p_{i_0}(t_0) = 0$. We claim that $p_i(t) = 0$ for all $t \in [0, +\infty[$ and $i \in I$. By the uniqueness of solutions it suffices to prove that $p_i(t_0) = 0$ for all $i \in I$. If there existed $i_1 \in I$ such that $p_{i_1}(t_0) > 0$, there would be $p_i(t_0) = 0$ and $\sum_{j \in \Omega_i} p_j(t_0) > 0$ for a suitable $i \in I$. Consequently,

$$p_i'(t_0) = \frac{\kappa}{\omega_i} \sum_{j \in \Omega_i} p_j(t_0) > 0.$$

This implies that p_i is strictly increasing on $]t_0 - \delta, t_0 + \delta[$ for a suitable $\delta > 0$, which leads to $p_i(t) < p_i(t_0) = 0$ for some $t \in]0, t_0[$, a contradiction. Likewise, $p_i(t) = 1$ for all $t \in [0, +\infty[$ and $i \in I$ if $p_{i_0}(t_0) = 1$. □

4. Asymptotic behaviour of the tumour model

Theorem 3. *If* $a \in P$ *and* $\inf_{i \in I} a_i > 0$, *then the solution* p *of* (2) *satisfies*

$$\lim_{t \to \infty} \sup_{i \in I} [1 - p_i(t)] = 0.$$

Accordingly, $\lim_{t \to +\infty} p_i(t) = 1$ *for all* $i \in I$.

Proof. Set $\rho = \inf_{i \in I} a_i > 0$. We claim that $p_i(t) \geq \rho$ for all $t \in [0, +\infty[$ and $i \in I$. We only need to show that $q_i(t) \geq \rho \sum_{m=0}^{n} \kappa^m t^m / m!$, for all $t \in [0, +\infty[$, $i \in I$, and $n \in N \cup \{0\}$, where $q_i(t) = \exp(\kappa t) p_i(t)$ $\forall t \in [0, +\infty[$.

From Theorem 2(ii) we deduce that

$$\begin{aligned} p_i'(t) &\geq -p_i(t) + \frac{\kappa}{\omega_i} \sum_{j \in \Omega_i} p_j(t) - (\kappa - 1) p_i(t) \\ &= \frac{\kappa}{\omega_i} \sum_{j \in \Omega_i} p_j(t) - \kappa p_i(t) \end{aligned}$$

and therefore

$$\begin{aligned} q_i'(t) &= \kappa \exp(\kappa t) p_i(t) + \exp(\kappa t) p_i'(t) \\ &\geq \kappa \exp(\kappa t) p_i(t) + \exp(\kappa t) \left(\frac{\kappa}{\omega_i} \sum_{j \in \Omega_i} p_j(t) - \kappa p_i(t) \right) \\ &= \frac{\kappa}{\omega_i} \sum_{j \in \Omega_i} q_j(t) \end{aligned}$$

for all $t \in [0, +\infty[$ and $i \in I$. In particular, $q_i'(t) \geq 0$ and therefore $q_i(t) \geq q_i(0) \geq \rho$ for all $t \in [0, +\infty[$ and $i \in I$. Assume the inequality holds for n; we will prove it for $n + 1$. Since $q_i(t) \geq \rho \sum_{m=0}^{n} \kappa^m t^m / m!$, we have

$$q_i'(t) \geq \frac{\kappa}{\omega_i} \sum_{j \in \Omega_i} q_j(t) \geq \frac{\kappa}{\omega_i} \sum_{j \in \Omega_i} \rho \sum_{m=0}^{n} \frac{\kappa^m t^m}{m!} = \rho \sum_{m=0}^{n} \frac{\kappa^{m+1} t^m}{m!}.$$

Thus

$$q_i(t) \geq q_i(0) + \rho \sum_{m=0}^{n} \frac{\kappa^{m+1} t^{m+1}}{(m+1)!} \geq \rho \sum_{m=0}^{n+1} \frac{\kappa^m t^m}{m!}$$

for all $t \in [0, +\infty[$ and $i \in I$, which proves our claim.

For every $n \in N$, let

$$\nu_n = \frac{\kappa^n \rho \prod_{m=1}^{n} (1 - (m+1)^{-2})}{1 + (\kappa^n - 1) \rho \prod_{m=1}^{n} (1 - (m+1)^{-2})}.$$

We next prove that for every $n \in N$ there is $t_n \in]0, +\infty[$ such that $t_n < t_{n+1}$ and $p_i(t) \geq \nu_n$ for all $i \in I$ and $t \in [t_n, +\infty[$. From what has already been proved, it follows that

$$\begin{aligned} p_i'(t) &= -p_i(t) + (\kappa - 1)(1 - p_i(t)) \frac{1}{\omega_i} \sum_{j \in \Omega_i} p_j(t) + \frac{1}{\omega_i} \sum_{j \in \Omega_i} p_j(t) \\ &\geq -p_i(t) + (\kappa - 1)(1 - p_i(t)) \rho + \rho \\ &= \kappa \rho - (1 + (\kappa - 1) \rho) p_i(t) \end{aligned}$$

for all $i \in I$ and $t \in [0, +\infty[$. For every $i \in I$ the function

$$u_i(t) = \exp((1 + (\kappa - 1)\rho)t)p_i(t)$$

satisfies

$$\begin{aligned} u_i'(t) &= (1 + (\kappa - 1)\rho)\mathrm{e}^{(1+(\kappa-1)\rho)t}p_i(t) + \mathrm{e}^{(1+(\kappa-1)\rho)t}p_i'(t) \\ &\geq (1+(\kappa - 1)\rho)\mathrm{e}^{(1+(\kappa-1)\rho)t}p_i(t) + \mathrm{e}^{(1+(\kappa-1)\rho)t}[\kappa\rho - (1 + (\kappa - 1)\rho)p_i(t)] \\ &\geq \kappa\rho\mathrm{e}^{(1+(\kappa-1)\rho)t}, \end{aligned}$$

hence

$$u_i(t) \geq p_i(0) + \frac{\kappa\rho}{1 + (\kappa - 1)\rho}(\mathrm{e}^{(1+(\kappa-1)\rho)t} - 1)$$

and therefore

$$\begin{aligned} p_i(t) &\geq p_i(0)\mathrm{e}^{-(1+(\kappa-1)\rho)t} + \frac{\kappa\rho}{1 + (\kappa - 1)\rho}(1 - \mathrm{e}^{-(1+(\kappa-1)\rho)t}) \\ &\geq \rho\mathrm{e}^{-(1+(\kappa-1)\rho)t} + \frac{\kappa\rho}{1 + (\kappa - 1)\rho}(1 - \mathrm{e}^{-(1+(\kappa-1)\rho)t}) \end{aligned}$$

for all $t \in [0, +\infty[$. Since

$$\lim_{t\to+\infty}\left[\rho\mathrm{e}^{-(1+(\kappa-1)\rho)t} + \frac{\kappa\rho}{1 + (\kappa - 1)\rho}(1 - \mathrm{e}^{-(1+(\kappa-1)\rho)t})\right] = \frac{\kappa\rho}{1 + (\kappa - 1)\rho}$$

and $\kappa\rho/(1 + (\kappa - 1)\rho) > \nu_1$, there exists $t_1 \in]0, +\infty[$ such that

$$\rho\mathrm{e}^{-(1+(\kappa-1)\rho)t} + \frac{\kappa\rho}{1 + (\kappa - 1)\rho}(1 - \mathrm{e}^{-(1+(\kappa-1)\rho)t}) \geq \nu_1$$

for all $t \in [t_1, +\infty[$. Consequently, $p_i(t) \geq \nu_1$ for all $i \in I$ and $t \in [t_1, +\infty[$. Assume that $t_n \in [0, +\infty[$ has been chosen satisfying our requirements. Then we have

$$\begin{aligned} p_i'(t) &= -p_i(t) + (\kappa - 1)(1 - p_i(t))\frac{1}{\omega_i}\sum_{j\in\Omega_i}p_j(t) + \frac{1}{\omega_i}\sum_{j\in\Omega_i}p_j(t) \\ &\geq -p_i(t) + (\kappa - 1)(1 - p_i(t))\nu_n + \nu_n \\ &= \kappa\nu_n - (1 + (\kappa - 1)\nu_n)p_i(t) \end{aligned}$$

for all $t \in [t_n, +\infty[$. Arguing as before we see that

$$p_i(t) \geq \nu_n\mathrm{e}^{-(1+(\kappa-1)\nu_n)(t-t_n)} + \frac{\kappa\nu_n}{1 + (\kappa - 1)\nu_n}(1 - \mathrm{e}^{-(1+(\kappa-1)\nu_n)(t-t_n)})$$

for all $i \in I$ and $t \in [t_n, +\infty[$. Since

$$\begin{aligned} &\lim_{t\to+\infty}\left[\nu_n\mathrm{e}^{-(1+(\kappa-1)\nu_n)(t-t_n)} + \frac{\kappa\nu_n}{1 + (\kappa - 1)\nu_n}(1 - \mathrm{e}^{-(1+(\kappa-1)\nu_n)(t-t_n)})\right] \\ &= \frac{\kappa\nu_n}{1 + (\kappa - 1)\nu_n} \end{aligned}$$

and

$$\frac{\kappa\nu_n}{1 + (\kappa - 1)\nu_n} = \frac{\kappa^{n+1}\rho\prod_{m=1}^{n}(1 - (m + 1)^{-2})}{1 + (\kappa^{n+1} - 1)\rho\prod_{m=1}^{n}(1 - (m + 1)^{-2})} > \nu_{n+1},$$

there exists $t_{n+1} \in]t_n, +\infty[$ such that

$$\nu_n e^{-(1+(\kappa-1)\nu_n)(t-t_n)} + \frac{\kappa\nu_n}{1+(\kappa-1)\nu_n}(1 - e^{-(1+(\kappa-1)\nu_n)(t-t_n)}) \geq \nu_{n+1}$$

for all $t \in [t_{n+1}, +\infty[$. Consequently, $p_i(t) \geq \nu_{n+1}$ for all $i \in I$ and $t \in [t_{n+1}, +\infty[$.

Since $\sup_{i\in I}[1 - p_i(t)] \leq 1 - \nu_n \; \forall t \in [t_n, +\infty[\; \forall n \in N$ and $\lim \nu_n = 1$, it may be concluded that $\lim_{t\to+\infty} \sup_{i\in I}[1 - p_i(t)] = 0$. □

COROLLARY 1

Assume that I is finite. If $a \in P$ and $a \neq 0$, then the solution p of (2) *satisfies*

$$\lim_{t\to+\infty} p_i(t) = 1$$

for all $i \in I$.

Proof. Let $t_0 \in]0, +\infty[$. Theorem 2(iii) now shows that $p_i(t_0) > 0 \; \forall i \in I$. The family $(q_i(t))_{i\in I} = (p_i(t + t_0))_{i\in I}$ is a solution of (2) with $a = (p_i(t_0))_{i\in I}$ which satisfies the requirement in the preceding theorem. Consequently, $\lim_{t\to\infty} q_i(t) = 1 \; \forall i \in I$, which proves the theorem. □

5. Mean tumour growth

In the remainder of this paper we assume that $a \in P$ and p is the solution of (2). For every $t \in [0, +\infty[$,

$$\mu(t) = \sum_{i\in I} p_i(t) \in [0, +\infty]$$

is the expected number of cancerous cells at the time t.

To study the function μ we introduce the set $l_1(I)$ set of all real families $x = (x_i)_{i\in I}$ such that $\|x\|_1 = \sum_{i\in I} |x_i| < +\infty$. $l_1(I)$ with pointwise operations and the norm $\|\cdot\|_1$ is a Banach space. It should be noted that $l_1(I) \subset l_\infty(I)$ and $\|x\|_\infty \leq \|x\|_1 \; \forall x \in l_1(I)$.

It is a simple matter to show that the restriction of f to $l_1(I)$ maps into $l_1(I)$. We write g for this restriction. If $a \in l_1(I)$ then we can consider the following initial value problem on $l_1(I)$

$$\begin{cases} x' &= g(x) \\ x(0) &= a. \end{cases} \tag{3}$$

Lemma 4. *The function g is infinitely differentiable on $l_1(I)$ and*

$$\|Dg(x)\| \leq 1 + 4\kappa + 5(\kappa - 1)\|x\|_\infty$$

for all $x \in l_1(I)$.

Proof. Let $x \in l_1(I)$ and S_x be the linear operator from $l_1(I)$ into itself defined by

$$E_i S_x(u) = -u_i + \frac{\kappa}{\omega_i}\sum_{j\in\Omega_i} u_j - (\kappa - 1)x_i\frac{1}{\omega_i}\sum_{j\in\Omega_i} u_j - (\kappa - 1)u_i\frac{1}{\omega_i}\sum_{j\in\Omega_i} x_j$$

for all $u \in l_1(I)$ and $i \in I$. We have

$$\begin{aligned}|E_iS_xu| &\leq |u_i| + \frac{\kappa}{\omega_i}\sum_{j\in\Omega_i}|u_j| + (\kappa-1)\|x\|_\infty\frac{1}{\omega_i}\sum_{j\in\Omega_i}|u_j| + (\kappa-1)|u_i|\|x\|_\infty \\ &\leq |u_i| + \kappa\sum_{j\in\Omega_i}|u_j| + (\kappa-1)\|x\|_\infty\sum_{j\in\Omega_i}|u_j| + (\kappa-1)|u_i|\|x\|_\infty\end{aligned}$$

for all $u \in l_1(I)$ and $i \in I$, and so

$$\|S_xu\|_1 \leq \|u\|_1 + 4\kappa\|u\|_1 + 4(\kappa-1)\|x\|_\infty\|u\|_1 + (\kappa-1)\|u\|_1\|x\|_\infty$$

for all $u \in l_1(I)$, since obviously $\sum_{i\in I}\sum_{j\in\Omega_i}|u_j| \leq 4\sum_{i\in I}|u_i|$. Consequently, S_x is continuous and $\|S_x\| \leq 1 + 4\kappa + 5(\kappa-1)\|x\|_\infty$. If $i \in I$ then

$$E_i(g(x+u) - g(x) - S_x(u)) = -(\kappa-1)u_i\frac{1}{\omega_i}\sum_{j\in\Omega_i}u_j$$

and therefore $\|g(x+u) - g(x) - S_x(u)\|_1 \leq 4(\kappa-1)\|u\|_1^2$ for all $u \in l_1(I)$. Thus g is differentiable at x with $Dg(x) = S_x$. From this g is easily checked to be 2 times differentiable on $l_1(I)$ with

$$E_i(D^2g(x)(u,v)) = -(\kappa-1)u_i\frac{1}{\omega_i}\sum_{j\in\Omega_i}v_j - (\kappa-1)v_i\frac{1}{\omega_i}\sum_{j\in\Omega_i}u_j$$

for all $x, u, v \in l_1(I)$. Since D^2g is constant, we conclude that $D^3g = 0$. □

Theorem 4. *If $\mu(0) < +\infty$ then $\mu(t) \leq \mu(0)e^{(5\kappa-4)t} < +\infty$ for all $t \in [0, +\infty[$.*

Proof. As $\mu(0) < +\infty$ we have $a \in P \cap l_1(I)$. From the preceding lemma and ([4]; 10.4.5 and 10.5.2) it follows that (3) has a unique noncontinuable solution q which is defined on an interval $]-\sigma, \tau[$ with $\sigma, \tau > 0$.

Since $\|\cdot\|_\infty \leq \|\cdot\|_1$, it follows immediately that q is differentiable on $]-\sigma,\tau[$ as a $l_\infty(I)$-valued function and obviously satisfies (2). Consequently, $q(t) = p(t)$ for all $t \in [0, \tau[$.

For every $t \in [0, \tau[$, the mean value theorem ([4]; 8.5.4) gives

$$\|g(q(t))\|_1 = \|g(q(t)) - g(0)\|_1 \leq \|q(t)\|_1 \sup_{0\leq\xi\leq1}\|Dg(\xi q(t))\|$$

and lemma 4 now shows that

$$\begin{aligned}\|g(q(t))\|_1 &\leq \|q(t)\|_1 \sup_{0\leq\xi\leq1}[1 + 4\kappa + 5(\kappa-1)\|\xi q(t)\|_\infty] \\ &\leq \|q(t)\|_1[1 + 4\kappa + 5(\kappa-1)].\end{aligned}$$

Since $q(t) = q(0) + \int_0^t g(q(s))ds$, we see that

$$\begin{aligned}\|q(t)\|_1 &\leq \|q(0)\|_1 + \int_0^t \|g(q(s))\|_1 ds \\ &\leq \|q(0)\|_1 + \int_0^t (5\kappa-4)\|q(s)\|_1 ds\end{aligned}$$

for all $t \in [0, \tau[$ and Gronwall lemma ([4]; 10.5.1.3) now yields

$$\|q(t)\|_1 \leq \|q(0)\|_1 \mathrm{e}^{(5\kappa-4)t}$$

for all $t \in [0, \tau[$. Therefore

$$\|q(t)\|_1 \leq \|q(0)\|_1 \mathrm{e}^{(5\kappa-4)\tau}$$

for all $t \in [0, \tau[$. On account of ([4]; 10.5.5 and 10.5.5.1), we have $\tau = +\infty$. Since $q(t) = p(t) \in [0, 1]$ $\forall t \in [0, +\infty[$, we conclude that $\mu(t) = \|q(t)\|_1 < +\infty$ for all $t \in [0, +\infty[$. □

Theorem 5. *If $\mu(0) < +\infty$ then the function μ is increasing and infinitely differentiable on $[0, +\infty[$ with $\mu^{(n)}(t) = \sum_{i\in I} p_i^{(n)}(t)$ for all $t \in [0, +\infty[$ and $n \in N$.*

Proof. In the proof of the preceding theorem we proved that p is the solution of (3), and consequently $p : [0, +\infty[\rightarrow l_1(I)$ is infinitely differentiable on $[0, +\infty[$. On the other hand, let φ be the continuous linear functional on $l_1(I)$ defined by $\varphi(x) = \sum_{i\in I} x_i$ $\forall x \in l_1(I)$. φ is infinitely differentiable on $l_1(I)$. Hence $\mu = \varphi \circ p$ is infinitely differentiable on $[0, +\infty[$ and an easy computation shows that $\mu^{(n)}(t) = \sum_{i\in I} p_i^{(n)}(t)$ for all $t \in [0, +\infty[$ and $n \in N$.

For every $t \in [0, +\infty[$, we have

$$\begin{aligned}
\mu'(t) &= \sum_{i\in I} p_i'(t) \\
&= -\sum_{i\in I} p_i(t) + \kappa \sum_{i\in I} \frac{1}{\omega_i}\left(\sum_{j\in\Omega_i} p_i(t)\right) - (\kappa - 1)\sum_{i\in I}\left(p_i(t)\frac{1}{\omega_i}\sum_{j\in\Omega_i} p_j(t)\right) \\
&= -\mu(t) + \kappa\mu(t) - (\kappa - 1)\sum_{i\in I}\left(p_i(t)\frac{1}{\omega_i}\sum_{j\in\Omega_i} p_j(t)\right) \\
&\geq -\mu(t) + \kappa\mu(t) - (\kappa - 1)\sum_{i\in I} p_i(t) = 0
\end{aligned}$$

which shows that μ is increasing on $[0, +\infty[$. □

From Corollary 1 and the preceding theorem we deduce the following.

COROLLARY 2

If I is finite then μ is an infinitely differentiable increasing function on $[0, +\infty[$. If $a \neq 0$, then $\lim_{t\rightarrow+\infty}\mu(t)$ equals the cardinality of I.

References

[1] Bramson M and Griffeath D, The asymptotic behavior of a probabilistic model for tumor growth, in: *Biological growth and spread.* Lecture Notes in Biomath. (Berlin: Springer-Verlag) (1979) vol. 38, pp. 165–172.

[2] Bramson M and Griffeath D, On the Williams–Bjerknes tumour growth model: II. *Math. Proc. Cambridge Philos. Soc.* **88** (1980) 339–357.

[3] Bramson M and Griffeath D, On the Williams–Bjerknes tumour growth model I. *Ann. Probab.* **9** (1981) 173–185.

[4] Dieudonne J, *Foundations of modern analysis.* (New York: Academic Press) (1960)

[5] Martinez F and Villena A R, An upper bound of the mean growth in the Williams–Bjerknes tumour model. *J. Appl. Anal.* **5** (1999) 277–282

[6] Richardson D, Random growth in a tessellation. *Math. Proc. Cambridge Philos. Soc.* **74** (1973) 515–528

[7] Schwartz D, Applications of duality to a class of Markov processes, *Ann. Probab.* **5** (1977) 522–532

[8] Williams T and Bjerknes R, Stochastic model for abnormal clone spread through epithelial basal layer. *Nature* **236** (1972) 19–21

Proc. Indian Acad. Sci. (Math. Sci.), Vol. 110, No. 3, August 2000, pp. 335–345
© Printed in India

Suppression of instability in rotatory hydromagnetic convection

JOGINDER S DHIMAN

Department of Mathematics, Himachal Pradesh University, Summer Hill, Shimla 171005, India

MS received 13 January 1999; revised 21 December 1999

Abstract. Recently discovered hydrodynamic instability [1], in a simple Bénard configuration in the parameter regime $T_0\alpha_2 > 1$ under the action of a nonadverse temperature gradient, is shown to be suppressed by the simultaneous action of a uniform rotation and a uniform magnetic field both acting parallel to gravity for oscillatory perturbations whenever $(\mathcal{Q}\sigma_1/\pi^2 + \mathcal{J}/\pi^4) > 1$ and the effective Rayleigh number $\mathcal{R}(1 - T_0\alpha_2)$ is dominated by either $27\pi^4(1 + 1/\sigma_1)/4$ or $27\pi^4/2$ according as $\sigma_1 \geq 1$ or $\sigma_1 \leq 1$ respectively. Here T_0 is the temperature of the lower boundary while α_2 is the coefficient of specific heat at constant volume due to temperature variation and $\sigma_1, \mathcal{R}, \mathcal{Q}$ and $\mathcal{J}$ respectively denote the magnetic Prandtl number, the Rayleigh number, the Chandrasekhar number and the Taylor number.

Keywords. Bénard convection; hydrodynamic instability; hydromagnetic instability; oscillatory perturbations.

1. Introduction

The study of thermal convection in a layer of fluid heated from below has many physical applications, notable in astrophysics and geophysics, as well as in the industrial processes. The problem of onset of convection in such a layer of fluid confined between two infinite horizontal planes, known as Bénard problem, has been extensively investigated both experimentally and theoretically. The Bénard problem was studied mathematically for the first time by Lord Rayleigh [9] for the idealized case of two free boundaries. Rayleigh's theory shows that the gravity-dominated thermal instability in liquid layer heated underside, depends upon the Rayleigh number which is proportional to the uniform temperature difference maintained between the lowermost and uppermost temperatures of the layers of the concerning liquid. Banerjee *et al* [1] presented a modified analysis of thermal instability of a liquid layer heated underside by emphasizing and utilizing the point, on the basis of the experimental results of Schmidt and Milverton [10], that linear theoretical explanation of the phenomenon of gravity-dominated thermal instability in a liquid layer heated underside should depend not only upon the Rayleigh number which is proportional to the uniform temperature difference maintained across the layer but also upon another parameter that takes care of a relatively hotter or a cooler layer under almost identical conditions. It was found that Rayleigh's utilization of the Boussinesq approximation overlooks a term in the equation of heat conduction which is on account of the variations in specific heat at constant volume due to the variations in temperature and which is such that in usual circumstances it cannot be ignored if the Boussinesq approximation were to be consistently and relatively more accurately applied throughout the analysis. The essential argument on which this term found a place in their modified theory is that it is the temperature differences that were of moderate amounts but not necessarily the temperature itself and an incorporation of this term into the calculations adequately completes the qualitative and quantitative gaps in Rayleigh's theory as pointed out earlier. Further they showed that the reformulated equations of the Bénard convection breakdown in the parameter regime $T_0\alpha_2 > 1$ and predict the existence of some new phenomenon. In particular the existence of the hydrodynamic instability in a single diffusive bottom heavy system is mathematically derived as an outcome of the reformulated equations in the parameter regime $T_0\alpha_2 > 1$.

The stability investigations of the Bénard problem in the framework of various external force fields assume importance not only on account of being a meaningful mathematical extension of the problem but also because of its importance in the problems of meteorology, oceanography and various other fields of practical importance. The effects of the action of a uniform magnetic field/a uniform rotation acting parallel to gravity on the Bénard problem has been investigated by Chandrasekhar [2] and others and it is shown that in some respect their individual/combined effects are remarkably alike, namely they both inhibit the onset of instability and elongate the cells which appear at the marginal stability for certain ranges of values of the parameters involved. Another interesting point brought out by Chandrasekhar's analysis which is, in general qualitative agreement with the experimental results of Nakagawa ([6,7] and [8]) Fultz, Nakagawa and Frenzen [4] and others is that, in both the problems the marginal state could either be stationary or oscillatory in character for which sufficient conditions are obtained. The work of Eltayeb [3] is also concerned with the combined effect of rotation and magnetic field on simple Bénard problem. Gupta *et al* [5] also investigated the problem under the joint influence of a rotation and a magnetic field especially with a view to derive bounds for the complex growth rate for an arbitrary osciallatory perturbation which may be neutral or unstable.

The aim of the present paper is to show how the instability reported by Banerjee *et al* is suppressed for oscillatory perturbations by the simultaneous application of a uniform vertical rotation and a uniform vertical magnetic field.

2. Construction of the modified simplified equations governing the problem

Let the origin be taken on the lower boundary with the positive direction of the z-axis along the vertically upward direction. Let $z = d\,(> 0)$ denote the upper boundary and T_0 and $T_1 (< T_0)$ respectively denote the uniform temperatures of the lower and upper boundaries. Further, let the layer of fluid be in a state of uniform rotation with angular velocity $\vec{\Omega}$ and subject to a uniform magnetic field $\vec{H}$ such that $\vec{\Omega}$ and $\vec{H}$ are parallel to gravity. Then following Banerjee–Gupta *et al* [1] the modified simplified equations governing the rotatory hydromagnetic Bénard convection problem are given by

$$\partial u_j/\partial x_j = 0, \tag{1}$$

$$\begin{aligned}\partial u_i/\partial t + u_j\partial u_i/\partial x_j - \mu_e|\vec{H}|^2(H_j\partial H_i/\partial x_j)/4\pi\rho_0 &= (1+\delta\rho/\rho_0)X_i\\ &- \partial\{p/\rho_0 - |\vec{\Omega}\times\vec{r}|^2/2 + \mu_e|\vec{H}|^2/8\pi\rho_0\}/\partial x_i + 2\epsilon_{ijk}u_j\Omega_k + \nu_0\nabla^2 u_i,\end{aligned} \tag{2}$$

$$(1 - T_0\alpha_2)\{\partial T/\partial t + \partial T/\partial x_j\} = K_0\nabla^2 T, \tag{3}$$

$$\partial H_i/\partial t + u_j\partial H_i/\partial x_j = H_j\partial u_i/\partial x_j + \eta_0\nabla^2 H_i, \tag{4}$$

$$\partial H_i/\partial x_j = 0 \tag{5}$$

and

$$\rho = \rho_0[1 - \alpha(T - T_0)], \tag{6}$$

where $x_j (j = 1, 2, 3)$ respectively denote the x, y and z coordinates; u_i, X_i, H_i, Ω_i $(i = 1, 2, 3)$ respectively denote the x, y and z components of velocity, external force, magnetic field and rotation; T denotes the temperature, ρ the density, p the pressure, ϵ_{ijk} the permutation tensor, μ_e the magnetic permeability, α is volume expansion, $\vec{r} = (x, y, z)$ is the position vector and ρ_0, ν_0, η_0 and K_0 stand for values of density, viscosity, magnetic diffusivity and thermal conductivity at the lower boundary $z = 0$.

Clearly, the initial stationary states whose stability we wish to examine is characterized by the following solutions for the velocity, temperature, magnetic field, density and pressure respectively:

$$u_j \equiv (0, 0, 0), T = T_0 - \beta z, H_i = (0, 0, H),$$
$$\rho = \rho_0[1 + \alpha(T_0 - T)] = \rho[1 + \alpha\beta z]$$

and

$$P = p - \rho_0|\vec{\Omega} \times \vec{r}|^2/2 + \mu_e|\vec{H}|^2/8\pi = P_0 - g\rho_0[z + \alpha\beta z^2/2], \tag{7}$$

where P_0 and ρ_0 are the values of P and ρ at the lower boundary $z = 0$ and $\beta = (T_0 - T_1)/d$ is the maintained uniform temperature gradient. Further, $(\Omega_i) = (0, 0, \Omega)$.

Let the initial stationary state described by equations (7) be slightly perturbed. Then the linearized perturbations equations on the basis of the normal mode resolution Chandrasekhar [2], wherein the desired solutions have x, y and t dependence of the form

$$\exp[i(k_x x + k_y y) + nt], \tag{8}$$

are as follows:

$$ik_x u + ik_y v + \mathrm{d}w/\mathrm{d}z = 0, \tag{9}$$

$$\rho_0 nu = -ik_x\delta P + \mu_0(\mathrm{d}^2/\mathrm{d}z^2 - k^2)u + (\mu_e H/4\pi)[\partial\hbar_x/\partial z - ik_x\hbar_z] + 2\rho_0\Omega v, \tag{10}$$

$$\rho_0 nv = -ik_y\delta P + \mu_0(\mathrm{d}^2/\mathrm{d}z^2 - k^2)v + (\mu_e H/4\pi)[\partial\hbar_y/\partial z - ik_y\hbar_z] - 2\rho_0\Omega u, \tag{11}$$

$$\rho_0 nw = -\mathrm{d}(\partial P)/\mathrm{d}z + \mu_0(\mathrm{d}^2/\mathrm{d}z^2 - k^2)w + g\alpha\rho_0\theta, \tag{12}$$

$$n(1 - T_0\alpha_2)\theta - (1 - T_0\alpha_2)\beta w = K_0(\mathrm{d}^2/\mathrm{d}z^2 - k^2)\theta, \tag{13}$$

$$n\hbar_x = H\mathrm{d}u/\mathrm{d}z + \eta_0(\mathrm{d}^2/\mathrm{d}z^2 - k^2)\hbar_x, \tag{14}$$

$$n\hbar_y = H\mathrm{d}v/\mathrm{d}z + \eta_0(\mathrm{d}^2/\mathrm{d}z^2 - k^2)\hbar_y, \tag{15}$$

$$n\hbar_z = H\mathrm{d}w/\mathrm{d}z + \eta_0(\mathrm{d}^2/\mathrm{d}z^2 - k^2)\hbar_z, \tag{16}$$

$$ik_x\hbar_x + ik_y\hbar_y + \mathrm{d}w/\mathrm{d}z = 0, \tag{17}$$

$$\rho_0 n\zeta = \mu_0(\mathrm{d}^2/\mathrm{d}z^2 - k^2)\zeta + (\mu_e H/4\pi)\mathrm{d}\xi/\mathrm{d}z + 2\rho_0\Omega\mathrm{d}w/\mathrm{d}z, \tag{18}$$

$$n\xi = H\mathrm{d}\zeta/\mathrm{d}z + \eta_0(\mathrm{d}^2/\mathrm{d}z^2 - k^2)\xi, \tag{19}$$

where $\{u(z), v(z), w(z)\}, \theta(z), \delta\rho(z), \delta P(z)$ and $\{\hbar_x(z), \hbar_y(z), \hbar_z(z)\}$ are the perturbations in the velocity, temperature, initial density, pressure and magnetic field respectively, $k = (k_x^2 + k_y^2)^{1/2}$ is the wave number of perturbation, k_x and k_y being real, n is a constant which can be complex in general, ζ and ξ denote the vorticity and the current density respectively.

Multiplying equation (10) by ik_x and equation (11) by ik_y, adding the resulting equations and making use of equations (9) and (17), we have

$$\begin{aligned}\rho_0 n\mathrm{d}w/\mathrm{d}z &= -k^2\delta P + \mu_0(\mathrm{d}^2/\mathrm{d}z^2 - k^2)\mathrm{d}w/\mathrm{d}z\\ &\quad + (\mu_e H/4\pi)(\mathrm{d}^2/\mathrm{d}z^2 - k^2)\hbar_z - 2\rho_0\Omega\zeta.\end{aligned}\tag{20}$$

Eliminating δP between eqs (20) and (12), it follows that

$$\begin{aligned}(\mathrm{d}^2/\mathrm{d}z^2 - k^2)(\mathrm{d}^2/\mathrm{d}z^2 - k^2 - n/\nu_0)w &= g\alpha k^2\theta/\nu_0\\ - (\mu_e H/4\pi\rho_0\nu_0)(\mathrm{d}^2/\mathrm{d}z^2 - k^2)\mathrm{d}\hbar_z/\mathrm{d}z &+ (2\Omega/\nu_0)\mathrm{d}\zeta/\mathrm{d}z.\end{aligned}\tag{21}$$

Further, eqs (13), (14)–(17), (18) and (19) can be written as

$$(\mathrm{d}^2/\mathrm{d}z^2 - k^2 - n/K_0)\theta = -(1 - T_0\alpha_2)\beta/K_0,\tag{22}$$

$$(\mathrm{d}^2/\mathrm{d}z^2 - k^2 - n/\eta_0)\hbar_z = -(H/\eta_0)\mathrm{d}w/\mathrm{d}z,\tag{23}$$

$$(\mathrm{d}^2/\mathrm{d}z^2 - k^2 - n/\nu_0)\zeta = -(\mu_e H/4\pi\rho_0\nu_0)\mathrm{d}\xi/\mathrm{d}z - (2\Omega/\nu_0)\mathrm{d}w/\mathrm{d}z,\tag{24}$$

$$(\mathrm{d}^2/\mathrm{d}z^2 - k^2 - n/\eta_0)\xi = -(H/\eta_0)\mathrm{d}\zeta/\mathrm{d}z.\tag{25}$$

Using the non-dimensional quantities defined by

$$\begin{aligned}&z_* = z/d;\ a_* = k/d;\ \sigma_* = \nu_0/K_0;\ D_* = d\mathrm{d}/\mathrm{d}z;\ \rho_* = nd/K_0^2;\ \sigma_{1*} = \nu_0/\eta_0;\\ &\theta_* = \mathcal{R}_* a^2\theta/\beta d;\ W_* = d(w/K_0);\ \xi_* = \nu_0\eta_0\xi/d(2\Omega HK_0);\ \hbar_{z*} = \eta_0\hbar_z/HK_0;\\ &\mathcal{R}_* = g\alpha\beta d^4/K_0\nu_0;\ \mathcal{Q}_* = \mu_e H^2 d^2/4\pi\rho_0\nu_0\eta_0;\ \mathcal{J}_* = 4\Omega^2 d^2/\nu_0.\end{aligned}\tag{26}$$

where $D \equiv \mathrm{d}/\mathrm{d}z, \nu_0 = \mu_0/\rho_0$ and dropping the asterisks for convenience in writing, we have the following system of equations

$$(D^2 - a^2)(D^2 - a^2 - p/\sigma)W = \theta - \mathcal{Q}D(D^2 - a^2)\hbar_z + \mathcal{J}D\zeta,\tag{27}$$

$$(D^2 - a^2 - p)\theta = -\mathcal{R}(1 - T_0\alpha_2)a^2 W,\tag{28}$$

$$(D^2 - a^2 - p\sigma_1\sigma)\hbar_z = -DW,\tag{29}$$

$$(D^2 - a^2 - p/\sigma)\zeta = -\mathcal{Q}D\xi - DW,\tag{30}$$

$$(D^2 - a^2 - p\sigma_1/\sigma)\xi = -D\zeta.\tag{31}$$

Equations (27)–(31) together with the boundary conditions

$$\begin{aligned}&W = 0 = \theta = DW = \zeta = D\xi = \hbar_z \quad \text{at} \quad z = 0 \quad \text{and} \quad z = 1,\\ &\text{(rigid boundaries with regions outside perfectly conducting)}\end{aligned}\tag{32}$$

constitute a double eigenvalue problem for p for prescribed values of $a^2, \mathcal{R}, \mathcal{Q}, \mathcal{J}$ and $T_0\alpha_2$ and a given state of the system is stable, neutral or unstable provided that the real part p_r of p is negative, zero or positive respectively. Further, if $p_r = 0$ implies that $p_i = 0$ for every wave number a, then the 'principle of exchange of stabilities' (PES) is valid, otherwise we have overstability at least when instability sets in as certain modes.

3. Mathematical analysis

We prove the following lemma and theorems;

Lemma. $(p, W, \theta, \hbar_z, \xi, \zeta)$, $p = p_r + ip_i$, $p_r \geq 0$, $p_i \neq 0$, $\mathcal{R} < 0$, $T_0\alpha_2 > 1$, $\mathcal{Q} > 0$, $\mathcal{J} > 0$ *is a solution of equations* (27)–(32), *then* $(\mathcal{Q}\sigma_1/\pi^2 + \mathcal{J}/\pi^4) > 1$.

Proof. Multiplying equation (27) by W^* (the complex conjugate of W) throughout, integrating the resulting equation over the vertical range of z and using eqs (28)–(31), we have

$$\begin{aligned}\int W^*(D^2 - a^2)(D^2 - a^2 - p/\sigma)W\mathrm{d}z &= -1/[\mathcal{R}(1 - T_0\alpha_2)]a^2 \\ &\times \int \theta(D^2 - a^2 - p^*)\theta^*\mathrm{d}z - \mathcal{Q}\int (D^2 - a^2)\hbar_z(D^2 - a^2 - p^*\sigma_1/\sigma)\hbar_z^*\mathrm{d}z \\ &+ \mathcal{J}\int \zeta(D^2 - a^2 - p^*/\sigma)\zeta^*\mathrm{d}z + \mathcal{Q}\mathcal{J}\int \xi^*(D^2 - a^2 - p\sigma_1/\sigma)\xi\mathrm{d}z. \qquad (33)\end{aligned}$$

The limits of integration in the above equation and subsequently will be omitted for sake of convenience in writing.

Integrating various terms of eq. (33) by parts for an appropriate number of times and making use of relevant boundary conditions given by eq. (32), we have

$$\begin{aligned}&\int (|D^2W|^2 + 2a^2|DW|^2 + a^4|W|^2)\mathrm{d}z + p/\sigma\int (|DW|^2 + a^2|W|^2)\mathrm{d}z \\ &+ \mathcal{J}\int (|D\zeta|^2 + a^2|\zeta|^2)\mathrm{d}z + \mathcal{J}p^*/\sigma\int |\zeta|^2\mathrm{d}z \\ &+ \mathcal{Q}\left[\int (|D^2\hbar_z|^2 + 2a^2|D\hbar_z|^2 + a^4|\hbar_z|^2)\mathrm{d}z + p^*\sigma_1/\sigma\int (|D\hbar_z|^2 + a^2|\hbar_z|^2)\mathrm{d}z\right] \\ &+ \mathcal{Q}\mathcal{J}\left[\int (|D\xi|^2 + a^2|\xi|^2 + (p\sigma_1/\sigma)|\xi|^2)\mathrm{d}z\right] \\ &= [1/[\mathcal{R}(1 - T_0\alpha_2)]a^2]\int [|D\theta|^2 + (a^2 + p^*)|\theta|^2]\mathrm{d}z. \qquad (34)\end{aligned}$$

Equating the real and imaginary parts of both sides of equation (34) and cancelling $p_i(\neq 0)$ throughout from the imaginary parts, we have

$$\begin{aligned}&\int (|D^2W|^2 + 2a^2|DW|^2 + a^4|W|^2)\mathrm{d}z + p_r/\sigma\int (|DW|^2 + a^2|W|^2)\mathrm{d}z \\ &+ \mathcal{J}\int (|D\zeta|^2 + a^2|\zeta|^2)\mathrm{d}z + \mathcal{Q}\int (|D^2\hbar_z|^2 + 2a^2|D\hbar_z|^2 + a^4|\hbar_z|^2)\mathrm{d}z \\ &+ \mathcal{Q}\mathcal{J}\int (|D\xi|^2 + a^2|\xi|^2)\mathrm{d}z + p_r/\sigma\left[\int (|D\hbar_z|^2 + a^2|\hbar_z|^2)\mathrm{d}z + \mathcal{J}\int |\zeta|^2\mathrm{d}z\right. \\ &\left.+ \mathcal{Q}\mathcal{J}\sigma_1\int |\xi|^2\mathrm{d}z - [\sigma/[\mathcal{R}(1 - T_0\alpha_2)]a^2]\int |\theta|^2\mathrm{d}z\right] \\ &= 1/[\mathcal{R}(1 - T_0\alpha_2)]a^2\int (|D\theta|^2 + a^2|\theta|^2)\mathrm{d}z \qquad (35)\end{aligned}$$

and

$$\begin{aligned}&1/\sigma\int (|DW|^2 + a^2|W|^2)\mathrm{d}z + \mathcal{Q}\mathcal{J}\sigma_1/\sigma\int |\xi|^2\mathrm{d}z + 1/[\mathcal{R}(1 - T_0\alpha_2)]a^2\int |\theta|^2\mathrm{d}z \\ &= \mathcal{Q}\sigma_1/\sigma\int (|D\hbar_z|^2 + a^2|\hbar_z|^2)\mathrm{d}z + \mathcal{J}/\sigma\int |\zeta|^2\mathrm{d}z. \qquad (36)\end{aligned}$$

Now multiplying both sides of equation (29) by $\hbar_z^*$, integrating the resulting equation over the vertical range of z a suitable number of times and making use of relevant boundary conditions (32), we have

$$\int(|D\hbar_z|^2 + a^2|\hbar_z|^2 + [p\sigma_1/\sigma]|\hbar_z|^2)\mathrm{d}z = -\int D\hbar_z^* W\mathrm{d}z. \tag{37}$$

Equating the real parts from both sides of the equation (37), we have

$$\begin{aligned}\int(|D\hbar_z|^2 + a^2|\hbar_z|^2 + [p_r\sigma_1/\sigma]|\hbar_z|^2)\mathrm{d}z &= \text{Real part of}\left[-\int D\hbar_z^* W\mathrm{d}z\right]\\ &\leq \left|-\int D\hbar_z^* W\mathrm{d}z\right| \leq \left|\int D\hbar_z^* W\mathrm{d}z\right| \leq \int |D\hbar_z^*||W|\mathrm{d}z = \int |D\hbar_z||W|\mathrm{d}z\\ &\leq (1/2)\left[\int |W|^2\mathrm{d}z + \int |D\hbar_z|^2\mathrm{d}z\right].\end{aligned} \tag{38}$$

Since, $p_r \geq 0$, therefore from inequality (38), we have

$$\int(|D\hbar_z|^2 + a^2|\hbar_z|^2)\mathrm{d}z < \int |W|^2\mathrm{d}z - a^2\int |\hbar_z|^2\mathrm{d}z < \int |W|^2\mathrm{d}z. \tag{39}$$

Further, since, $W(0) = 0 = W(1)$, we have the Rayleigh–Ritz inequality [10]

$$\int |W|^2\mathrm{d}z \leq 1/\pi^2 \int |DW|^2\mathrm{d}z. \tag{40}$$

Multiplying both sides of eq. (30) by ζ^*, integrating the resulting equation over the vertical range of z a suitable number of times and making use of relevant boundary conditions (32), we have

$$\int(|D\zeta|^2 + a^2|\zeta|^2 + [p/\sigma]|\zeta|^2)\mathrm{d}z = Q\int \zeta^* D\xi\mathrm{d}z + \int \zeta^* DW\mathrm{d}z. \tag{41}$$

Equating the real parts from both sides of the equation (41), we have

$$\begin{aligned}\int(|D\zeta|^2 + a^2|\zeta|^2 + p_r/\sigma|\zeta|^2)\mathrm{d}z &= \text{Real part of}\left(Q\int \zeta^* D\xi\mathrm{d}z + \int \zeta^* DW\mathrm{d}z\right)\\ &= \text{Real part of}\left(Q\int \zeta^* D\xi\mathrm{d}z - \int D\zeta^* W\,\mathrm{d}z\right).\end{aligned} \tag{42}$$

Also

$$\int \zeta^* D\xi\mathrm{d}z = -\int \xi D\zeta^*\mathrm{d}z). \tag{43}$$

Now, substituting the value of $D\zeta^*$ from eq. (31) in eq. (43), integrating the resulting equation over the vertical range of z a suitable number of times and using the relevant boundary conditions (32), we have

$$\int \zeta^* D\xi\mathrm{d}z = -\int(|D\xi|^2 + a^2|\xi|^2 + [p^*\sigma_1/\sigma]|\xi|^2)\mathrm{d}z.$$

Therefore,

$$\text{Real part of} \int \zeta^* D\xi\mathrm{d}z = -\int(|D\xi|^2 + a^2|\xi|^2 + [p_r\sigma_1/\sigma]|\xi|^2)\mathrm{d}z. \tag{44}$$

Since, $p_r \geq 0$, it follows from eq. (44) that

$$\text{Real part of } \int \zeta^* D\xi \mathrm{d}z < 0. \tag{45}$$

Consequently, eq. (42) implies that

$$\begin{aligned}&\int (|D\zeta|^2 + a^2|\zeta|^2 + p_r/\sigma|\zeta|^2)\mathrm{d}z < \text{Real part of}\left(-\int D\zeta^* W\,\mathrm{d}z\right)\\ &\leq \left|\int D\zeta^* W \mathrm{d}z\right| \leq \int |D\zeta^*||W|\mathrm{d}z = \int |D\zeta||W|\mathrm{d}z < \left[\int |W|^2\mathrm{d}z\right]^{1/2}\\ &\times \left[\int |D\zeta|^2\mathrm{d}z\right]^{1/2} \text{(using Schwartz-inequality).}\end{aligned} \tag{46}$$

Since, $p_r \geq 0$, inequality (46) implies that $\int |D\zeta|^2\mathrm{d}z < \int |W|^2\mathrm{d}z$ which upon using Rayleigh–Ritz inequality $\int |D\zeta|^2\mathrm{d}z \geq \pi^2 \int |\zeta|^2\mathrm{d}z$, gives

$$\int |\zeta|^2\mathrm{d}z < 1/\pi^2 \int |W|^2\mathrm{d}z. \tag{47}$$

Using inequalities (39) (40) and (47) in eq. (36), we get

$$\begin{aligned}&[\pi^2/\sigma - \{\mathcal{Q}\sigma_1/\sigma + \mathcal{J}/\sigma\pi^2\}] \int |W|^2\mathrm{d}z + a^2/\sigma \int |W|^2\mathrm{d}z\\ &+\frac{\mathcal{Q}\mathcal{J}\sigma_1}{\sigma} \int |\xi|^2\mathrm{d}z + 1/[\mathcal{R}(1 - T_0\alpha_2)]a^2 \int |\theta|^2\mathrm{d}z < 0.\end{aligned} \tag{48}$$

It clearly follows from inequality (48) that $(\mathcal{Q}\sigma_1/\pi^2 + \mathcal{J}/\pi^4) > 1$. This completes the proof.

The above lemma proves that the oscillatory modes $(p_i \neq 0)$ of the problem under consideration will be stable $(p_r < 0)$ when $(\mathcal{Q}\sigma_1/\pi^2 + \mathcal{J}/\pi^4) \leq 1$, or equivalently, a necessary condition for the existence of oscillatory modes which may be neutral $(p_r = 0)$ or unstable $(p_r > 0)$ for the problem under consideration is that $(\mathcal{Q}\sigma_1/\pi^2 + \mathcal{J}/\pi^4) > 1$.

Theorem 1. $(p, W, \theta, \hbar_z, \xi, \zeta), p = p_r + ip_i, p_i \neq 0, (\mathcal{Q}\sigma_1/\pi^2 + \mathcal{J}/\pi^4) > 1, \mathcal{R} < 0, T_0\alpha_2 > 1$, $\mathcal{Q} > 0$, $\mathcal{J} > 0$ *is a solution of eqs* (27)–(32), *and* $\sigma_1 \geq 1$, $\mathcal{R}[1 - T_0\alpha_2] \leq 27\pi^4(1 + 1/\sigma_1)/4$, *then* $p_r < 0$.

Proof. Proceeding exactly as in the proof of lemma, we get eqs (35) and (36). If permissible, let $p_r \geq 0$. Multiplying eq. (36) by p_r and adding the resulting equation to eq. (35) we have

$$\begin{aligned}&\int (|D^2W|^2 + 2a^2|DW|^2 + a^4|W|^2)\mathrm{d}z\\ &+2p_r/\sigma\left[\int (|DW|^2 + a^2|W|^2)\mathrm{d}z + \mathcal{Q}\mathcal{J}\sigma_1 \int |\xi|^2\mathrm{d}z\right]\\ &+\mathcal{J}\int (|D\zeta|^2 + a^2|\zeta|^2)\mathrm{d}z + \mathcal{Q}\int (|D^2\hbar_z|^2 + 2a^2|D\hbar_z|^2 + a^4|\hbar_z|^2)\mathrm{d}z\\ &+\mathcal{Q}\mathcal{J}\int (|D\xi|^2 + a^2|\xi|^2)\mathrm{d}z = 1/[\mathcal{R}(1 - T_0\alpha_2)]a^2 \int (|D\theta|^2 + a^2|\theta|^2)\mathrm{d}z.\end{aligned} \tag{49}$$

Multiplying eq. (28) by its complex conjugate and integrating over the range of z by parts an appropriate number of times, using the boundary conditions (32) and equating the real parts of the resulting equation, we obtain

$$\begin{aligned}&\int (|D^2\theta|^2 + 2a^2|D\theta|^2 + a^4|\theta|^2)\mathrm{d}z + 2p_r \int (|D\theta|^2 + a^2|\theta|^2)\mathrm{d}z \\ &\quad \times |p|^2 \int |\theta|^2 \mathrm{d}z = [\mathcal{R}(1 - T_0\alpha_2)]^2 a^4 \int |W|^2 \mathrm{d}z. \end{aligned} \tag{50}$$

Since, $p_r \geq 0$, therefore eq. (50) implies that

$$\begin{aligned}&\int (|D^2\theta|^2 + 2a^2|D\theta|^2 + a^4|\theta|^2)\mathrm{d}z \\ &\quad = \int |(D^2 - a^2)\theta|^2 \mathrm{d}z < [\mathcal{R}(1 - T_0\alpha_2)]^2 a^4 \int |W|^2 \mathrm{d}z. \end{aligned} \tag{51}$$

Further

$$\begin{aligned}&\int (|D\theta|^2 + a^2|\theta|^2)\mathrm{d}z = \left| - \int \theta^*(D^2 - a^2)\theta \mathrm{d}z \right| \leq \left| \int \theta^*(D^2 - a^2)\theta \mathrm{d}z \right| \\ &\quad \leq \int |\theta||(D^2 - a^2)\theta|\mathrm{d}z \leq \left[\int |\theta|^2 \mathrm{d}z\right]^{1/2} \times \left[\int |(D^2 - a^2)\theta|^2 \mathrm{d}z\right]^{1/2} \\ &\quad\quad \text{(using Schwartz-inequality)} \end{aligned} \tag{52}$$

and

$$\begin{aligned}&\int |D\theta|^2 \mathrm{d}z = \left| - \int \theta^* D^2\theta \mathrm{d}z \right| \leq \left| \int \theta^* D^2\theta \mathrm{d}z \right| \leq \int |\theta^*||D^2\theta|\mathrm{d}z \\ &\quad = \int |\theta||D^2\theta|\mathrm{d}z \leq \left[\int |\theta|^2 \mathrm{d}z\right]^{1/2} \times \left[\int |D^2\theta|^2 \mathrm{d}z\right]^{1/2} \\ &\quad\quad \text{(using Schwartz-inequality).} \end{aligned} \tag{53}$$

Since, $\theta(0) = 0 = \theta(1)$, using Rayleigh–Ritz inequality, we have

$$\pi^2 \int |\theta|^2 \mathrm{d}z \leq \int |D\theta|^2 \mathrm{d}z. \tag{54}$$

Using inequality (53), inequality (52) implies that

$$\pi^4 \int |\theta|^2 \mathrm{d}z \leq \int |D^2\theta|^2 \mathrm{d}z. \tag{55}$$

Now combining inequalities (53) and (54), we have

$$\begin{aligned}&\int (|D^2\theta|^2 + 2a^2|D\theta|^2 + a^4|\theta|^2)\mathrm{d}z \geq \int (\pi^4|\theta|^2 + 2a^2\pi^2|\theta|^2 + a^4|\theta|^2)\mathrm{d}z \\ &\quad = (\pi^2 + a^2)^2 \int |\theta|^2 \mathrm{d}z, \end{aligned} \tag{56}$$

which combined with inequality (51) yields the inequality

$$(\pi^2 + a^2)^2 \int (|\theta|^2)\mathrm{d}z < [\mathcal{R}(1 - T_0\alpha_2)]^2 a^4 \int |W|^2 \mathrm{d}z. \tag{57}$$

Hence, inequality (52) with the help of inequalities (51) and (57) gives

$$\int(|D\theta|^2 + a^2|\theta|^2)\mathrm{d}z < \mathcal{R}[(1 - T_0\alpha_2)]^2 a^4/(\pi^2 + a^2)\int |W|^2\mathrm{d}z. \tag{58}$$

Further, eq. (36) implies that

$$\begin{aligned}\mathcal{Q}\sigma_1\int(|D\hbar_z|^2 + a^2|\hbar_z|^2)\mathrm{d}z + \mathcal{J}\int|\zeta|^2\mathrm{d}z &> \int(|DW|^2 + a^2|W|^2)\mathrm{d}z\\ &> (\pi^2 + a^2)\int|W|^2\mathrm{d}z\end{aligned}$$

or

$$\mathcal{Q}\int(|D\hbar_z|^2 + a^2|\hbar_z|^2)\mathrm{d}z > (\pi^2 + a^2)/\sigma_1\int|W|^2)\mathrm{d}z - \mathcal{J}/\sigma_1\int|\zeta|^2\mathrm{d}z. \tag{59}$$

Also, we have the following inequalities which are derived in a manner analogous to the derivations of inequalities (54)–(56) (since $W(0) = 0 = W(1)$, $\hbar_z(0) = 0 = \hbar_z(1)$ and $\zeta(0) = 0 = \zeta(1)$):

$$\int(|D^2W|^2 + 2a^2|DW|^2 + a^4|W|^2)\mathrm{d}z \geq (\pi^2 + a^2)^2\int|W|^2\mathrm{d}z,$$

$$\int(|D^2\hbar_z|^2 + 2a^2|D\hbar_z|^2 + a^4|\hbar_z|^2)\mathrm{d}z \geq (\pi^2 + a^2)\int(|D\hbar_z|^2 + a^2|\hbar_z|^2)\mathrm{d}z$$

and

$$\int(|D\zeta|^2 + a^2|\zeta|^2)\mathrm{d}z \geq (\pi^2 + a^2)\int|\zeta|^2\mathrm{d}z. \tag{60}$$

Equation (49) upon using inequalities (58)–(60) yields the following inequality

$$\begin{aligned}&[(\pi^2 + a^2)^3/a^2 + (\pi^2 + a^2)^3/a^2\sigma_1 - [\mathcal{R}(1 - T_0\alpha_2)]]\int|W|^2\mathrm{d}z\\ &\quad + \mathcal{J}(\pi^2 + a^2)^2(1 - 1/\sigma_1)/a^2\int|\zeta|^2\mathrm{d}z + 2p_r(\pi^2 + a^2)/a^2\sigma\\ &\qquad \times\left[\int(|DW|^2 + a^2|W|^2)\mathrm{d}z + \mathcal{Q}\mathcal{J}\sigma_1\int|\xi|^2\mathrm{d}z\right]\\ &\quad + \mathcal{Q}\mathcal{J}(\pi^2 + a^2)/a^2\int(|D\xi|^2 + a^2|\xi|^2)\mathrm{d}z < 0.\end{aligned} \tag{61}$$

Now, since the minimum value of $(\pi^2 + a^2)^3/a^2$ with respect to a^2 is $27\pi^4/4$, therefore it follows from inequality (61) that

$$\begin{aligned}&[27\pi^4(1 + 1/\sigma_1)/4 - [\mathcal{R}(1 - T_0\alpha_2)]]\int|W|^2\mathrm{d}z\\ &\quad + \mathcal{J}(\pi^2 + a^2)^2(1 - 1/\sigma_1)/a^2\int|\zeta|^2\mathrm{d}z\\ &\quad + 2p_r(\pi^2 + a^2)/a^2\sigma\left[\int(|DW|^2 + a^2|W|^2)\mathrm{d}z + \mathcal{Q}\mathcal{J}\sigma_1\int|\xi|^2\mathrm{d}z\right]\\ &\quad + \mathcal{Q}\mathcal{J}(\pi^2 + a^2)/a^2\int(|D\xi|^2 + a^2|\xi|^2)\mathrm{d}z < 0\end{aligned} \tag{62}$$

which clearly is incompatible with the hypothesis of Theorem 1. Hence, if $\sigma_1 \geq 1$ and $[\mathcal{R}(1 - T_0\alpha_2)] \leq 27\pi^4(1 + 1/\sigma_1)/4$ then $p_r < 0$. This completes the proof.

Theorem 2. *$(p, W, \theta, \hbar_z, \xi, \zeta)$, $p = p_r + ip_i$, $p_i \neq 0$, $(\mathcal{Q}\sigma_1/\pi^2 + \mathcal{J}/\pi^4) > 1$, $\mathcal{R} < 0$, $T_0\alpha_2 > 1$, $\mathcal{Q} > 0$, $\mathcal{J} > 0$ is a solution of eqs (27)–(32), and $\sigma_1 \leq 1$, $\mathcal{R}[1 - T_0\alpha_2] \leq 27\pi^4/2$, then $p_r < 0$.*

Proof. Proceeding exactly as in the proof of the Theorem 1, eq. (36) can be written as

$$\begin{aligned}\mathcal{J}\int|\zeta|^2\mathrm{d}z &> \int(|DW|^2 + a^2|W|^2)\mathrm{d}z - \mathcal{Q}\sigma_1\int(|D\hbar_z|^2 + a^2|\hbar_z|^2)\mathrm{d}z\\ &> (\pi^2 + a^2)\int|W|^2\mathrm{d}z - \mathcal{Q}\sigma_1\int(|D\hbar_z|^2 + a^2|\hbar_z|^2)\mathrm{d}z. \qquad (63)\end{aligned}$$

Now, using inequality (63) in place of inequality (59) and inequalities (58) and (60) in eq. (49), we have

$$\begin{aligned}&[(\pi^2 + a^2)^3/a^2 + (\pi^2 + a^2)^3/a^2 - [\mathcal{R}(1 - T_0\alpha_2)]]\int|W|^2\mathrm{d}z\\ &\quad + \mathcal{Q}(\pi^2 + a^2)^2(1 - \sigma_1/a^2)\int(|D\hbar_z|^2 + a^2|\hbar_z|^2)\mathrm{d}z\\ &\quad + 2p_r(\pi^2 + a^2)/a^2\sigma\left[\int(|DW|^2 + a^2|W|^2)\mathrm{d}z + \mathcal{Q}\mathcal{J}\sigma_1\int|\xi|^2\mathrm{d}z\right]\\ &\quad + \mathcal{Q}\mathcal{J}(\pi^2 + a^2)/a^2\int(|D\xi|^2 + a^2|\xi|^2)\mathrm{d}z < 0, \qquad (64)\end{aligned}$$

which upon substituting the minimum value of $(\pi^2 + a^2)^3/a^2$ with respect to a^2, i.e. $27\pi^4/4$, reduces to

$$\begin{aligned}&[2(27\pi^4/4) - [\mathcal{R}(1 - T_0\alpha_2)]]\int|W|^2\mathrm{d}z\\ &\quad + \mathcal{Q}(\pi^2 + a^2)^2(1 - \sigma_1)/a^2\int(|D\hbar_z|^2 + a^2|\hbar_z|^2)\mathrm{d}z\\ &\quad + 2p_r(\pi^2 + a^2)/a^2\sigma\left[\int(|DW|^2 + a^2|W|^2)\mathrm{d}z + \mathcal{Q}\mathcal{J}\sigma_1\int|\xi|^2\mathrm{d}z\right]\\ &\quad + \mathcal{Q}\mathcal{J}(\pi^2 + a^2)/a^2\int(|D\xi|^2 + a^2|\xi|^2)\mathrm{d}z < 0. \qquad (65)\end{aligned}$$

Inequality (65) is clearly incompatible with the hypothesis of Theorem 2. Hence, if $\sigma_1 \leq 1$ and $\mathcal{R}[1 - T_0\alpha_2] \leq 27\pi^4/2$, then $p_r < 0$. This completes the proof.

4. Conclusion

Theorems 1 and 2 show that the oscillatory modes of the rotatory magnetohydrodynamic modified Bénard convection problem can be suppressed by the simultaneous application of a uniform vertical rotation and a uniform vertical magnetic field in the parameter regime $\mathcal{R}[1 - T_0\alpha_2] \leq 27\pi^4(1 + 1/\sigma_1)/4$ or $27\pi^4/2$ according as $\sigma_1 \geq 1$ or $\sigma_1 \leq 1$.

Acknowledgement

Thanks are extended to the learned referee for his perspicacious comments on an earlier version of the paper.

References

[1] Banerjee M B, Gupta J R, Shandil R G and Jyoti Prakash, Breakdown of classical equations and existence of hydrodynamic instability in single diffusive bottom heavy system, *J. Math. Anal. Appl.* **173** (1993) 458

[2] Chandrasekhar S, *Hydrodynamic and Hydromagnetic Stability*, (London: Oxford University Press, Amen. House) (1961) E.C. 4

[3] Eltayeb I A, Overstable hydromagnetic convection in a rotating fluid layer, *J. Fluid Mech.* **71** (1975) 161

[4] Fultz D, Nakagawa Y and Frenzen P C, An instance in thermal convection of Eddington's 'Overstability', *Phys. Rev.* **94** (1954) 1471

[5] Gupta J R, Sood S K and Bhardwaj U D, On Rayleigh-Bénard convection with rotation and magnetic field, *J. Appl. Math. Phys.* **35** (1984) 252

[6] Nakagawa Y, An experiment on the inhibition of thermal convection by a magnetic field, *Nature*, **175** (1955) 417

[7] Nakagawa Y, Experiments on the inhibition of thermal convection by a magnetic field, *Proc. R. Soc. London* **A240** (1957) 108

[8] Nakagawa Y, Experiments on the instability of a layer of mercury heated from below and subject to the simultaneous action of a magnetic field and rotation-II, *Proc. R. Soc. London* **A249** (1959) 138

[9] Rayleigh L, On the convective currents in a horizontal layer of fluid when the higher temperature is on the upper side, *Philos. Mag.* **3** (1916) 529

[10] Schmidt R J and Milverton S W, On the instability of the fluid when heated from below, *Proc. R. Soc. London*, **A152** (1935) 586

[11] Schultz M H, Spline analysis (NJ: Eaglewood Prentice Hall, Cliffs) (1973)

Proc. Indian Acad. Sci. (Math. Sci.), Vol. 110, No. 4, November 2000, pp. 347–356.
© Printed in India

A quantum spin system with random interactions I

STEPHEN DIAS BARRETO

Stat. Math. Unit, Indian Statistical Institute, Bangalore 560 059, India
Current address: Department of Mathematics, Goa University, Sub P.O. Goa University, Taleigao Plateau, Goa 403 206
Email: girish@unigoa.ernet.in

MS received 9 February 2000; revised 17 April 2000

Abstract. We study a quantum spin glass as a quantum spin system with random interactions and establish the existence of a family of evolution groups $\{\tau_t(\omega)\}_{\omega\in\Omega}$ of the spin system. The notion of ergodicity of a measure preserving group of automorphisms of the probability space Ω, is used to prove the almost sure independence of the Arveson spectrum $\mathrm{Sp}(\tau(\omega))$ of $\tau_t(\omega)$. As a consequence, for any family of $(\tau(\omega),\beta)$-KMS states $\{\rho(\omega)\}$, the spectrum of the generator of the group of unitaries which implement $\tau(\omega)$ in the GNS representation is also almost surely independent of ω.

Keywords. Spin; system; quasi-local; random; dynamics; evolution; independent; Arveson; KMS.

1. Introduction

Traditionally, spin glasses [1] have been studied as spin systems with random interactions. These models are essentially Ising-type models with random coupling. Extensive investigations on the existence of the thermodynamic limit have been made e.g. van Enter *et al* [5, 6], and the equilibrium statistical mechanics of such systems has been studied. In order to study the dynamics of a quantum spin glass we model it as a quantum spin system on an infinite lattice with random interactions. We introduce a measure preserving group of automorphisms $\{T_a\}_{a\in\mathbb{Z}^\nu}$ with an ergodic action on a complete probability space $(\Omega, \mathcal{S}, P)$, and consider the class of interactions Φ which satisfy a compatibility condition involving the measure preserving group of automorphisms and the action of the lattice $\mathbb{Z}^\nu$ on the quasi-local algebra. We establish the existence of a family $\{\tau_t(\omega)\}$, of strongly continuous one-parameter groups of *-automorphisms of the quasi-local algebra $\mathcal{A}$ associated with the infinite system. Some interesting algebraic properties of $\tau_t(\omega)$ as well as those of its generator have been derived. Finally, we show that the Arveson spectrum of the evolution group $\tau_t(\omega)$ is almost surely independent of ω. Also, for any given family of $(\tau(\omega),\beta)$-KMS states $\{\rho(\omega)\}$, we report an ergodic property of the spectrum of the generator of the group of unitaries which implement $\tau(\omega)$ in the GNS representation. Results pertaining to the ergodic properties of the spectra of random self adjoint operators have also been reported in chapter 1 of [8]. However the techniques used there differ from our approach.

2. Description of the random model

We consider a quantum spin-$\frac{1}{2}$ system with spins located at the vertices of an infinite lattice $\mathbb{Z}^\nu$. The interaction between spins of course is taken to be random. The kinematical structure of the spin system is described by a quasi-local (UHF) algebra $\mathcal{A}$ with a generating net of $C^\star$-subalgebras $\{\mathcal{A}_\Lambda\}$, constructed over the finite subsets Λ of $\mathbb{Z}^\nu$. Besides this, we also have a natural action α of the lattice $\mathbb{Z}^\nu$ as *-automorphisms $a \mapsto \alpha_a$ of $\mathcal{A}$. For a detailed description see § 6.2.4 in [10] and § 6.2.1 in [4].

3. Random interactions

Before introducing random interactions, one defines the notion of measurability of Banach space valued functions on a measure space $(\Omega, \mathcal{S}, m)$, where Ω is a set, $\mathcal{S}$ a sigma algebra and m a sigma-finite measure on Ω.

DEFINITION 3.1

Let $(\Omega, \mathcal{S}, m)$ be a measure space. A function $f : \Omega \to B$ where B is a Banach space, is said to be weakly measurable if, for every $\phi \in B^\star$, the map $\omega \mapsto \phi(f(\omega))$ is $\mathcal{S}$-measurable. f is said to be strongly measurable if, there exists a sequence of countably valued functions strongly convergent to f almost everywhere on Ω [7].

In case m is a finite measure, then we may replace 'countably valued' in the above definition by 'simple'. It can be shown that the notions of strong and weak measurability are equivalent if B is separable.

From now on, let $(\Omega, \mathcal{S}, P)$ be a complete probability space.

DEFINITION 3.2

Let $\mathcal{F}$ be the collection of all finite subsets of $\mathbb{Z}^\nu$. A random interaction is a map $\Phi : \mathcal{F} \times \Omega \to \mathcal{A}$ such that, for each $\omega \in \Omega$, $\Phi(X, \omega)$ is a self-adjoint element in the $C^\star$-subalgebra $\mathcal{A}_X$ and $\omega \mapsto \Phi(X, \omega)$ is strongly measurable for every $X \in \mathcal{F}$.

Now given the random interaction Φ, for finite $\Lambda \subseteq \mathbb{Z}^\nu$, the Hamiltonian associated with the spins confined to the region Λ is given by a Hermitian (self adjoint) element

$$H(\Lambda, \omega) = \sum_{X \subseteq \Lambda} \Phi(X, \omega),$$

for each realization $\omega \in \Omega$. Clearly, $H(\Lambda, \omega)$ is strongly measurable since each $\Phi(X, \omega)$ is strongly measurable on Ω.

Next, we introduce a measure preserving group of automorphisms $\{T_a\}_{a \in \mathbb{Z}^\nu}$ on the probability space Ω, and consider the class of only those random interactions Φ which satisfy the following condition:

$$\Phi(X + a, T_{-a}\omega) = \alpha_a(\Phi(X, \omega)),$$

see [5]. Clearly, $H(\Lambda + a, T_{-a}\omega) = \alpha_a(H(\Lambda, \omega)) \forall \omega \in \Omega$ and $a \in \mathbb{Z}^\nu$.

DEFINITION 3.3

Let Φ be a random interaction. Then Φ is said to be a finite range interaction if, the set

$$\Delta_\omega = \{x \in \mathbb{Z}^\nu | \exists X \ni x; \text{ such that } 0 \in X, \text{ and } \Phi(X, T_a\omega) \neq 0, \text{ for some } a \in \mathbb{Z}^\nu\}$$

is a finite subset of $\mathbb{Z}^\nu$ for almost every $\omega \in \Omega$.

Remark. Clearly, whenever $X - X \not\subseteq \Delta_\omega$, $\Phi(X, \omega) = 0$.

The definition given above yields the following result.

Lemma 3.4. *Let Φ be a random interaction. Then $\Delta_\omega = \Delta_{T_b\omega}$ for all $b \in Z^\nu$.*

Proof. Proof follows from the definition of Δ_ω. △

4. Random evolution

For a finite spin system confined to a region $\Lambda \subseteq \mathbb{Z}^\nu$, and for $\omega \in \Omega$, the equation of motion is given by

$$\frac{dA_t^\Lambda(\omega)}{dt} = i[H(\Lambda,\omega), A_t^\Lambda(\omega)], \quad A_t^\Lambda(\omega) \in \mathcal{A}_\Lambda.$$

This yields the time evolution given by $\tau_t^\Lambda(\omega)(A) = A_t^\Lambda(\omega) = e^{iH(\Lambda,\omega)t}Ae^{-iH(\Lambda,\omega)t}$ for $\omega \in \Omega$ and for all $A \in \mathcal{A}_\Lambda$. Clearly, for each $\omega \in \Omega$, $\tau_t^\Lambda(\omega)$ is a one-parameter group of *-automorphisms of $\mathcal{A}_\Lambda$. Since the spin system consists of infinite number of spins, the construction of the time evolution of a fixed observable $A \in \mathcal{A}_{\Lambda_0}$ where $\Lambda_0 \subseteq \mathbb{Z}^\nu$, involves taking the limit of $\tau_t^\Lambda(\omega)(A)$ as $\Lambda \to \infty$ (the collection $\mathcal{F}$ of all finite subsets Λ of $\mathbb{Z}^\nu$ ordered by inclusion is a directed set).

Next we construct a family of strongly continuous one-parameter groups of *-automorphisms which determine the evolution of the spin system. To this end, we have the following theorem.

Theorem 4.1. *Let Φ be a finite range random interaction of the quantum spin system on a lattice $\mathbb{Z}^\nu$, satisfying*

$$\sup_{a\in\mathbb{Z}^\nu}\left(\sum_{X\ni 0}||\Phi(X,T_a\omega)||\right) < \infty$$

almost everywhere. Then, for almost every $\omega \in \Omega$, there exists a strongly continuous, one-parameter group of *-automorphisms $\tau_t(\omega)$ of $\mathcal{A}$ such that,

$$\lim_{\Lambda\to\infty} \tau_t^\Lambda(\omega)(A) = \tau_t(\omega)(A), \quad \forall\, A \in \mathcal{A}$$

and uniformly, for t in compacts, where $\tau_t^\Lambda(\omega)(A) = e^{iH(\Lambda,\omega)t}A\, e^{-iH(\Lambda,\omega)t}$. $\tau_t(\omega)$ is called the evolution group of the spin system whenever the limit exists.

Proof. Since $\Phi(X+a, T_{-a}\omega) = \alpha_a(\Phi(X,\omega))$, $\forall a \in \mathbb{Z}^\nu$, it is clear that whenever Δ_ω is finite,

$$P_\phi(\omega)(x) = \sum_{X\ni x}||\Phi(X,\omega)|| = \sum_{Y\ni 0}||\Phi(Y,T_x\omega)|| < \infty.$$

On appealing to Proposition 6.2.3 in [4], there exists a derivation $\delta(\omega)$ of $\mathcal{A}$ such that, the domain of $\delta(\omega)$

$$D(\delta(\omega)) = \bigcup_{\Lambda\subseteq\mathbb{Z}^\nu} \mathcal{A}_\Lambda; \quad \delta(\omega)(A) = i\sum_{X\cap\Lambda\neq\emptyset}[\Phi(X,\omega),A] \quad \text{for } A \in \mathcal{A}_\Lambda,$$

and $\delta(\omega)$ is norm-closeable with norm closure $\overline{\delta}(\omega)$. Next, we shall show that $D(\delta(\omega))$ is a dense set of analytic elements for $\overline{\delta}(\omega)$. Take $A \in \mathcal{A}_{\Lambda_0}$. Whenever Δ_ω is a finite set and

$$\sup_{a\in\mathbb{Z}^\nu}\left(\sum_{X\ni 0}||\Phi(X,T_a\omega)||\right)<\infty,$$

we have

$$(\overline{\delta}(\omega))^n(A)=i^n\sum_{X_1\cap S_0\neq\emptyset}\cdots\sum_{X_n\cap S_{n-1}\neq\emptyset}[\Phi(X_n,\omega),[\dots[\Phi(X_1,\omega),A]]],$$

where

$$S_0=\Lambda_0 \quad\text{and}\quad S_j=\Lambda_0\cup\bigcup_{i=1}^{j}X_i,\quad \text{for}\quad j\geq 1.$$

Now, if

$$[\Phi(X_j,\omega),[\dots[\Phi(X_1,\omega),A]]]\neq 0,$$

then

$$|X_i|\leq|\Delta_\omega|,\quad \forall\, i=1,2,\dots,j \quad\text{and therefore,}\quad |S_j|\leq j|\Delta_\omega|+|\Lambda_0|.$$

Here $|\cdot|$ denotes the cardinality of a set. Therefore we get

$$\begin{aligned}||(\overline{\delta}(\omega))^n(A)||&\leq 2^n||A||\sum_{x_1\in S_0}\sum_{X_1\ni x_1}\cdots\sum_{x_n\in S_{n-1}}\sum_{X_n\ni x_n}||\Phi(X_n,\omega)||\cdots||\Phi(X_1,\omega)||\\&\leq 2^n||A||\sum_{x_1\in S_0}\sum_{X_1-x_1\ni 0}\cdots\sum_{x_n\in S_{n-1}}\sum_{X_n-x_n\ni 0}||\Phi(X_n-x_n,T_{x_n}\omega)||\\&\qquad\cdots||\Phi(X_1-x_1,T_{x_1}\omega)||\\&\leq 2^n||A||\prod_{i=1}^{n}((i-1)|\Delta_\omega|+|\Lambda_0|)\left(\sup_{a\in\mathbb{Z}^\nu}\left(\sum_{X\ni 0}||\Phi(X,T_a\omega)||\right)\right)^n\\&\leq ||A||\mathrm{e}^{|\Lambda_0|}2^n n!\left(\sup_{a\in\mathbb{Z}^\nu}\left(\sum_{X\ni 0}||\Phi(X,T_a\omega)||\right)\right)^n \mathrm{e}^{n|\Delta_\omega|}.\end{aligned}$$

This establishes that A is an analytic element for $\overline{\delta}(\omega)$, [2] with radius of analyticity independent of A. Therefore, it follows from Proposition 6.2.3 in [4] and the assumptions made in this theorem that for almost every $\omega\in\Omega$, $\overline{\delta}(\omega)$ is the generator of a strongly continuous one parameter group of *-automorphisms $\tau_t(\omega)$ of $\mathcal{A}$ such that,

$$\tau_t^\Lambda(\omega)(A)\to\tau_t(\omega)(A),\quad \forall\, A\in\mathcal{A},$$

where convergence above is uniform in t on compact sets. △

From now on let $\mu\times P$ be the complete product measure on $\mathbb{R}\times\Omega$, μ being the Lebesque measure on $\mathbb{R}$.

PROPOSITION 4.2

Let $\tau_t(\omega)$ be the strongly continuous, one-parameter group of *-automorphisms of $\mathcal{A}$, constructed above. Then, $(t,\omega)\mapsto\tau_t(\omega)(A)$ is strongly, jointly measurable in t and ω, for all $A\in\mathcal{A}$.

Proof. The proof is a consequence of theorem 4.1, if one notices that there exists a sequence $\{\Lambda_n\}$ of finite subsets increasing to $\mathbb{Z}^\nu$ i.e.,

$$\Lambda_1 \subset \Lambda_2 \subset \Lambda_3 \subset \cdots, \quad \text{and} \quad \bigcup_{n=1}^{\infty} \Lambda_n = \mathbb{Z}^\nu. \qquad \triangle$$

It is seen in the case of quantum spin systems on a lattice $\mathbb{Z}^\nu$ with translation invariant interactions that, whenever the dynamics exists, the evolution group of *-automorphisms of the quasi-local algebra commutes with the action of the lattice $\mathbb{Z}^\nu$ on the algebra. Here we prove a variant of this property. Before we set about establishing this result the following fact is worth noting.

Lemma 4.3. *Let $\tau_t^\Lambda(\omega)$ be the one-parameter group of *-automorphisms associated with a finite $\Lambda \subseteq \mathbb{Z}^\nu$, where*

$$\tau_t^\Lambda(\omega)(A) = \mathrm{e}^{iH(\Lambda,\omega)t} A \mathrm{e}^{-iH(\Lambda,\omega)t}, \quad \forall\, A \in \mathcal{A}.$$

Then for all $a \in \mathbb{Z}^\nu$, we have

$$\alpha_a(\tau_t^\Lambda(\omega)(A)) = \tau_t^{\Lambda+a}(T_{-a}\omega)(\alpha_a(A)); \quad \forall A \in \mathcal{A}.$$

Proof. Using the fact that α_a is a *-automorphism, the lemma follows from functional calculus for $H(\Lambda,\omega)$, and the identity $H(\Lambda + a, T_{-a}\omega) = \alpha_a(H(\Lambda,\omega))$. △

PROPOSITION 4.4

Let $\tau_t(\omega)$ be the evolution group of the spin system on an infinite lattice $\mathbb{Z}^\nu$. Then for all $a \in \mathbb{Z}^\nu$, we have

$$\tau_t(T_{-a}\omega)(\alpha_a(A)) = \alpha_a(\tau_t(\omega)(A)), \quad \forall\, A \in \mathcal{A}.$$

Proof. The proof follows from theorem 4.1, and the lemma established prior to this proposition. △

In the proposition that follows, we establish an interesting algebraic property of the generators $\overline{\delta}(\omega)$ of the evolution groups $\tau_t(\omega)$.

PROPOSITION 4.5

Let $\tau_t(\omega)$ be the evolution group of the spin system and $D(\overline{\delta}(\omega))$ be the domain of its generator. Then for all $a \in \mathbb{Z}^\nu$, we have

$$\alpha_a(D(\overline{\delta}(\omega))) = D(\overline{\delta}(T_{-a}\omega)) \quad \textit{and} \quad \alpha_a(\overline{\delta}(\omega))(A) = \overline{\delta}(T_{-a}\omega)(\alpha_a(A)),$$

for all $A \in D(\overline{\delta}(\omega))$.

Proof. Throughout the proof of this theorem $\delta^\Lambda(\omega)$ will denote the generator $i[H(\Lambda,\omega),.]$ of $\tau_t^\Lambda(\omega)$. Let $\{\Lambda_n\}$ be a sequence of finite subsets increasing to $\mathbb{Z}^\nu$. Since $\tau_t^{\Lambda_n}(\omega)(B) \to \tau_t(\omega)(B)$; $\forall B \in \mathcal{A}$, we conclude from preliminary 2.4 in [2] that, $\overline{\delta}(\omega)$ is the graph limit of $\delta^{\Lambda_n}(\omega)$. Hence, for $A \in D(\overline{\delta}(\omega))$, there exists a sequence $\{A_n\}$, where $A_n \in D(\delta^{\Lambda_n}(\omega))$ such that, $A_n \to A$ and $\delta^{\Lambda_n}(\omega)(A_n) \to \overline{\delta}(\omega)(A)$. This implies that $\alpha_a(A_n) \to \alpha_a(A)$ and

$\alpha_a(\delta^{\Lambda_n}(\omega)(A_n)) \to \alpha_a(\bar{\delta}(\omega)(A))$. Now, since $\alpha_a(\delta^{\Lambda_n}(\omega)(A_n)) = \delta^{\Lambda_n+a}(T_{-a}\omega)(\alpha_a(A_n))$, we have $\delta^{\Lambda_n+a}(T_{-a}\omega)(\alpha_a(A_n)) \to \alpha_a(\bar{\delta}(\omega)(A))$. Since $\tau_t^{\Lambda_n+a}(T_{-a}\omega)(B)$ converges to $\tau_t(T_{-a}\omega)$ (B), for all $B \in \mathcal{A}$, preliminary 2.4 in [2] implies that $\bar{\delta}(T_{-a}\omega)$ is the graph limit of $\delta^{\Lambda_n+a}(T_{-a}\omega)$. Therefore, one concludes that $\alpha_a(A) \in D(\bar{\delta}(T_{-a}\omega))$ and $\alpha_a(\bar{\delta}(\omega))(A) = \bar{\delta}(T_{-a}\omega)(\alpha_a(A))$. Conversely, it can be shown that if $A \in D(\bar{\delta}(T_{-a}\omega))$ then $\alpha_{-a}(A) \in D(\bar{\delta}(\omega))$. This completes the proof of the proposition. △

In the next section, we study the Arveson spectrum of the evolution group $\tau_t(\omega)$, and report an interesting ergodic property of the Arveson spectrum.

5. Arveson spectrum

Here we introduce the notion of Arveson spectrum. If $\mathcal{A}$ is a C^*-algebra and $t \mapsto \gamma_t$, $t \in \mathbb{R}$, a strongly continuous, one-parameter group of *-automorphisms of the C^*-algebra, then the Bochner integral

$$\int_{-\infty}^{\infty} f(t)\gamma_t(A)\mathrm{d}t = \Gamma(f)(A); \quad A \in \mathcal{A}, \quad f \in L^1(\mathbb{R}),$$

defines a representation of $L^1(\mathbb{R})$ into the bounded operators on $\mathcal{A}$. Next, we have the following definition.

DEFINITION 5.1

The Arveson spectrum $\mathrm{Sp}(\gamma)$ of γ is a subset of the dual group $\hat{\mathbb{R}}$ of $\mathbb{R}$ defined as

$$\mathrm{Sp}(\gamma) = \{\sigma \in \mathbb{R} | \hat{f}(\sigma) = 0, \quad \forall f \in \ker \Gamma\},$$

where $\hat{f}$ is the Fourier transform of f.

It can be shown that $s \in \mathrm{Sp}(\gamma)$, if and only if, $|\hat{f}(s)| \le ||\Gamma(f)||$, for all $f \in L^1(\mathbb{R})$ vide [9].

The following definition is in order.

DEFINITION 5.2

Let $(\Omega, \mathcal{S}, P)$ be a probability space and J some index set. If T_j is a measure preserving automorphism of Ω, for each $j \in J$, then the action of T_j's is said to be ergodic if, for $A \in \mathcal{S}$, $P(A) = 0$ or 1 whenever $T_jA = A$, for all $j \in J$.

Our aim is to show that the Arveson spectrum of the evolution group $\tau_t(\omega)$ is almost surely independent of ω. To this end, we have the following theorem.

Theorem 5.3. *Let $\tau_t(\omega)$ be the strongly continuous, one-parameter group of *-automorphisms of $\mathcal{A}$, which determines the evolution of the spin system. If the action of the measure preserving group of automorphisms $\{T_a\}$ is ergodic, then the Arveson spectrum $Sp(\tau(\omega))$ of $\tau_t(\omega)$ is almost surely independent of ω.*

Proof. For $s \in \mathbb{R}$, let $E_s = \{\omega \in \Omega : ||\Gamma(\omega)(f)|| \ge |\hat{f}(s)| \forall f \in L^1(\mathbb{R})\}$, where

$$\Gamma(\omega)(f)(A) = \int_{-\infty}^{\infty} f(t)\tau_t(\omega)(A)\mathrm{d}t, \quad \forall A \in \mathcal{A}.$$

We show that E_s is a measurable subset of Ω. Since $L^1(\mathbb{R})$ is separable, there exists a countable dense set $F = \{f_n | n = 1, 2, \ldots\}$ in $L^1(\mathbb{R})$. Hence, for each $f \in L^1(\mathbb{R})$, there exists a sequence f_{n_k} in F, converging to f in the L^1-norm. It follows from the properties of the Bochner integral that

$$\begin{aligned} | \, ||\Gamma(\omega)(f_{n_k})|| - ||\Gamma(\omega)(f)|| \, | &\le ||\Gamma(\omega)(f_{n_k}) - \Gamma(\omega)(f)|| \\ &= ||\Gamma(\omega)(f_{n_k} - f)|| \\ &= \sup_{||A||=1} ||\int_{-\infty}^{\infty} (f_{n_k} - f)(t)\tau_t(\omega)(A)\mathrm{d}t|| \\ &\le \sup_{||A||=1} \left(||A|| \int_{-\infty}^{\infty} |(f_{n_k} - f)(t)|\mathrm{d}t \right) \\ &= \int_{-\infty}^{\infty} |(f_{n_k} - f)(t)|\mathrm{d}t \\ &= ||f_{n_k} - f||_1 . \end{aligned}$$

Therefore, $||\Gamma(\omega)(f_{n_k})||$ converges to $||\Gamma(\omega)(f)||$, for f_{n_k} converging to f, in the L^1-norm. In view of this, and the fact that F is dense in $L^1(\mathbb{R})$, we have

$$E_s = \bigcap_{n=1}^{\infty} E_s^n,$$

where $E_s^n = \{\omega \in \Omega | \, ||\Gamma(\omega)(f_n)|| \ge |\hat{f}_n(s)|\}$. In order to show that each of these E_s^n's is a measurable subset of Ω, it is sufficient to establish the measurability of the function $\omega \mapsto ||\Gamma(\omega)(f_n))||$, for all $n = 1, 2, \ldots$. On appealing to Proposition 4.2, we conclude that for $f \in L^1(\mathbb{R})$ and $A \in \mathcal{A}$, $(t, \omega) \mapsto f(t)\tau_t(\omega)(A)$ is strongly, jointly measurable in t and ω. Moreover,

$$\begin{aligned} \int_{\mathbb{R}\times\Omega} ||f(t)\tau_t(\omega)(A)||\mathrm{d}(\mu \times P)(t, \omega) &= \int_{\mathbb{R}\times\Omega} |f(t)| \, ||\tau_t(\omega)(A)||\mathrm{d}(\mu \times P)(t, \omega) \\ &= \int_{\mathbb{R}} \int_{\Omega} ||A|| \, |f(t)|\mathrm{d}\mu(t)\mathrm{d}P(\omega) < \infty. \end{aligned}$$

Hence, it follows that $(t, \omega) \mapsto f(t)\tau_t(\omega)(A)$ is Bochner integrable on $\mathbb{R} \times \Omega$ [7]. Therefore, as a consequence of the analogue of Fubini's theorem for vector valued functions (see [7]), the map $\omega \mapsto \Gamma(\omega)(f)(A)$ is strongly measurable in ω. Hence, $\omega \mapsto ||\Gamma(\omega)(f)(A)||$ is a measurable, real valued function on Ω. Thus it readily follows that for $f \in L^1(\mathbb{R})$, $\omega \mapsto ||\Gamma(\omega)(f)(A)||$ is measurable for all $A \in \mathcal{A}$. Now $\mathcal{A}$ being a separable $C^\star$-algebra, we have for $c \ge 0$ and $f \in L^1(\mathbb{R})$,

$$\{\omega \in \Omega | \, ||\Gamma(\omega)(f)|| \le c\} = \bigcap_{n=1}^{\infty} \{\omega \in \Omega | \, ||\Gamma(\omega)(f)(A_n)|| \le c; ||A_n|| \le 1\},$$

where $\mathcal{U}_0 = \{A_n \in \mathcal{A} | n = 1, 2, \ldots\}$ is a dense subset of the closed unit ball in $\mathcal{A}$. This, coupled with the fact that $\omega \mapsto ||\Gamma(\omega)(f)(A_n)||$ is a measurable function of ω for all $n = 1, 2, \ldots$, permits us to conclude that the set $\{\omega \in \Omega | \, ||\Gamma(\omega)(f)|| \le c\}$, is a measurable subset of Ω. Since c is arbitrary, the function $\omega \mapsto ||\Gamma(\omega)(f)||$ is a measurable function of ω. Thus, $\omega \mapsto ||\Gamma(\omega)(f)||$ is measurable for all $f \in L^1(\mathbb{R})$. Therefore, $\omega \mapsto ||\Gamma(\omega)f_n||$ is measurable $\forall n = 1, 2, \ldots$. Hence, each of these E_s^n's is a measurable subset of Ω. This proves conclusively that the set E_s is a measurable subset of Ω. Now, using the fact that the action of the measure preserving group of automorphisms $\{T_a\}$ is ergodic, we show that set E_s has measure either zero or one. It follows from the properties of the Bochner integral [10] and the fact that α_a is a *-automorphism of the C^*-algebra $\mathcal{A}$ that, for $f \in L^1(\mathbb{R})$,

$$\begin{aligned}
||\Gamma(\omega)(f)|| &= \sup_{||A||=1} \left|\left| \int_{-\infty}^{\infty} f(t)\tau_t(\omega)(A)\mathrm{d}t \right|\right| \\
&= \sup_{||\alpha_a(A)||=1} \left|\left| \alpha_a\left(\int_{-\infty}^{\infty} f(t)\tau_t(\omega)(A)\mathrm{d}t \right) \right|\right| \\
&= \sup_{||\alpha_a(A)||=1} \left|\left| \int_{-\infty}^{\infty} f(t)\tau_t(T_{-a}\omega)(\alpha_a(A))\mathrm{d}t \right|\right| \\
&= ||\Gamma(T_{-a}\omega)(f)||,
\end{aligned}$$

for all $a \in \mathbb{Z}^\nu$. The penultimate step follows from Proposition 4.4. Therefore it is clear from the above equalities, that the set E_s is invariant under the action of the measure preserving group of automorphisms $\{T_a\}$. Since the action of the measure preserving group of automorphisms $\{T_a\}$ is assumed to be ergodic, it follows that the set E_s has measure either zero or one. Hence, s lies in the Arveson spectrum of $\tau_t(\omega)$ with probability either zero or one. Thus, the Arveson spectrum $\mathrm{Sp}(\tau(\omega))$ of $\tau_t(\omega)$ is almost surely independent of ω. △

DEFINITION 5.4

Let $(\mathcal{A}, \tau)$ be a C^*-dynamical system, ρ a state over $\mathcal{A}$. Then for $\beta > 0$, ρ is said to be a (τ, β)-KMS state if, for any pair $A, B \in \mathcal{A}$, there exists a complex function $F_{A,B}$ which is analytic on the open strip $0 < \Im z < \beta$, uniformly bounded and continuous on the closed strip $0 \leq \Im z \leq \beta$ such that, $F_{A,B}(t) = \rho(A\tau_t(B))$ and $F_{A,B}(t + i\beta) = \rho(\tau_t(B)A)$.

Next, we shall show that for any family of $(\tau(\omega), \beta)$-KMS states the spectrum of the generator of the unitary group $U_t(\omega)$, which implements $\tau(\omega)$ in the GNS representation is almost surely independent of ω. To this end, we have the following proposition.

PROPOSITION 5.5

Let $\{\rho(\omega)\}$ be a family of $(\tau(\omega), \beta)$-KMS states. Also, let $H(\omega)$ be the generator of the strongly continuous, one-parameter group of unitaries $U_t(\omega)$ which implement $\tau_t(\omega)$ in the GNS representation. Then the spectrum $\sigma(H(\omega))$ of the generator $H(\omega)$ is almost surely independent of ω.

Proof. Let π_ω denote the representation associated with the $(\tau(\omega), \beta)$-KMS state $\rho(\omega)$, with cyclic vector Θ_ω. The unitary group $U_t(\omega)$ with generator $H(\omega)$ implements $\tau_t(\omega)$ in this representation π_ω. Now, for $f \in L^1(\mathbb{R})$, we have

$$\Psi_\omega(f)\phi = \int_{-\infty}^{\infty} f(t)U_t(\omega)\phi\mathrm{d}t = 0, \quad \forall\ \phi \in \mathcal{H}_\omega$$

$$\Leftrightarrow \int_{-\infty}^{\infty} f(t)U_t(\omega)(\pi_\omega(A)\Theta_\omega)\mathrm{d}t = 0, \quad \forall\, A \in \mathcal{A}$$

$$\begin{aligned}
&\Leftrightarrow \int_{-\infty}^{\infty} f(t)\pi_\omega(\tau_t(\omega)(A))\Theta_\omega \mathrm{d}t = 0, \quad \forall\, A \in \mathcal{A}\\
&\Leftrightarrow \left(\int_{-\infty}^{\infty} f(t)\pi_\omega(\tau_t(\omega)(A))\mathrm{d}t\right)\Theta_\omega = 0, \quad \forall\, A \in \mathcal{A}\\
&\Leftrightarrow \pi_\omega\left(\int_{-\infty}^{\infty} f(t)\tau_t(\omega)(A)\mathrm{d}t\right)\Theta_\omega = 0, \quad \forall\, A \in \mathcal{A}\\
&\Leftrightarrow \pi_\omega\left(\int_{-\infty}^{\infty} f(t)\tau_t(\omega)(A)\mathrm{d}t\right) = 0, \quad \forall\, A \in \mathcal{A}\\
&\Leftrightarrow \int_{-\infty}^{\infty} f(t)\tau_t(\omega)(A)\mathrm{d}t = 0, \quad \forall\, A \in \mathcal{A}.
\end{aligned}$$

The first step follows from the fact that Θ_ω is a cyclic vector for $\pi_\omega(\mathcal{A})$. The second follows from the definition of $U_t(\omega)$. Since $\rho(\omega)$ is a KMS state, the separating character of the cyclic vector Θ_ω for $\pi_\omega(\mathcal{A})''$, accounts for the penultimate step. We arrive at the final step by virtue of the fact that the representation π_ω is faithful. Now, using the spectral theorem it is not very difficult to show that $\sigma(H(\omega)) = -\{s \in \mathbb{R} | \hat{f}(s) = 0, \forall f \in \ker \Psi(\omega)\}$. Therefore, we have $\sigma(H(\omega)) = -\{s \in \mathbb{R} | \hat{f}(s) = 0,\ \forall f \in \ker \Gamma(\omega)\}$, where $\Gamma(\omega)$ is as in the theorem proved above. Hence, the proof follows from the theorem proved above, where we have shown that the Arveson spectrum of $\tau(\omega)$ is almost surely independent of ω. △

6. Conclusion

In this paper we have studied the dynamics of a quantum spin glass through the spectral properties of a family of evolution groups $\{\tau_t(\omega)\}$ of a quantum spin system with random interactions. The almost sure independence of the Arveson spectrum $\mathrm{Sp}(\tau(\omega))$ of the evolution group $\tau_t(\omega)$ in a way suggests that the Arveson spectrum becomes independent of ω in the thermodynamic limit. Besides, given a family of $(\tau(\omega), \beta)$-KMS states $\{\rho(\omega)\}$, we demonstrated the almost sure independence of the spectrum of the generator of the group of unitaries which implement the evolution group $\tau(\omega)$ in the GNS representation.

Acknowledgements

This work is dedicated to the memory of my father, late Dr Antonio Dias Barreto. It is a pleasure to thank Prof. Y S Prahalad for suggesting the problem, and for a number of useful conversations. I also thank the referee for drawing my attention to results on spectra of random self-adjoint operators reported in [8]. This research was supported by the Council for Scientific and Industrial Research.

References

[1] Binder K and Young A P, Spin glasses: Experimental facts, theoretical concepts and open questions, *Rev. Mod. Phys.* **58** (1986) 801–976

[2] Bratteli O and Kishimoto A, Generation of semi-groups and two dimensional quantum lattice systems, *J. Funct. Anal.* **35** (1980) 344–368

[3] Bratteli O and Robinson D W, *Operator algebras and quantum statistical mechanics* (New York: Springer-Verlag) (1979) vol. 1

[4] Bratteli O and Robinson D W, *Operator algebras and quantum statistical mechanics* (New York: Springer-Verlag) (1981) vol. 2

[5] van Enter A C D and van Hemmen J L, The thermodynamic limit for long-range random systems, *J. Stat. Phys.* **32** (1983) 141–152

[6] van Enter A C D and van Hemmen J L, Statistical–mechanical formalism for spin glasses, *Phys. Rev.* **A29** (1984) 355–365

[7] Hille E and Phillips R, *Functional analysis and semi-groups*, revised edition (American Mathematical Society Colloquium Publications, Providence, RI) (1957) vol. 31

[8] Pastur L and Figotin A, *Spectra of random and almost-periodic operators* (Berlin: Springer-Verlag) (1992)

[9] Pedersen G, *$C^\star$-algebras and their automorphism groups* (New York: Academic Press) (1979)

[10] Ruelle D, *Statistical mechanics* (New York: W A Benjamin Inc.) (1969)

Proc. Indian Acad. Sci. (Math. Sci.), Vol. 110, No. 4, November 2000, pp. 357–362.
© Printed in India

Weighted approximation of continuous functions by sequences of linear positive operators

TÜLIN COŞKUN

Department of Mathematics, Karaelmas University, 67100 Zonguldak, Turkey
E-mail: tcoskun@karaelmas.edu.tr

MS received 24 November 1999; revised 7 July 2000

Abstract. In this work we obtain, under suitable conditions, theorems of Korovkin type for spaces with different weight, composed of continuous functions defined on unbounded regions. These results can be seen as an extension of theorems by Gadjiev in [4] and [5].

Keywords. Korovkin theorem; positive linear operators; weighted spaces; weight functions.

Let $C(a,b)$ denote the space of all continuous functions on $[a,b]$ and let $B(a,b)$ be the space of all bounded functions on the same interval. If the sequence of positive linear operators $A_n : C(a,b) \longrightarrow B(a,b)$ satisfy the three conditions

$$\lim_{n\to\infty} \|A_n(1,x) - 1\|_{C(a,b)} = 0,$$

$$\lim_{n\to\infty} \|A_n(t,x) - x\|_{C(a,b)} = 0,$$

$$\lim_{n\to\infty} \|A_n(t^2,x) - x^2\|_{C(a,b)} = 0,$$

then

$$\lim_{n\to\infty} \|A_n(f,x) - f(x)\|_{C(a,b)} = 0$$

for all function $f \in C(a,b)$ for which $|f(x)| \leq M_f(1+x^2)$ hold on $\mathbb{R}$. This theorem is known as Korovkin theorem ([6, 1]) and it is important in approximation theory. The theorem shows that convergence on three functions may be extended to all functions which are continuous on $[a,b]$ and bounded on $\mathbb{R}$. Baskakov [2] generalized this result to unbounded functions on $\mathbb{R}$.

In refs [4] and [5] Gadjiev defined the weight spaces C_ρ and B_ρ of real functions defined on the real line and showed that Korovkin's theorem in general does not hold on these spaces. Here $B_\rho := \{f : |f| \leq M_f \cdot \rho,\ \rho \geq 1 \text{ and } \rho \text{ unbounded}\}$ and $C_\rho := \{f : f \in B_\rho \text{ and } f \text{ continuous}\}$ are spaces of functions which are defined on unbounded sets. However in [4] and [5] it has been shown that this theorem holds on a common subspace of the spaces B_ρ and C_ρ.

In ref. [3] it is proved that a theorem of Korovkin type does not hold on the spaces C_{ρ_1} and B_{ρ_2} with different weights ρ_1 and ρ_2, respectively. But in this study we show that if we put some appropriate conditions on the weight functions it holds.

We firstly give the following lemma which will be needed for proving the other theorems.

Let

$$\psi_n(s) := \sup_{\|f\|_{\rho_1}=1} \sup_{|x|\leq s} \frac{|A_n(f,x)|}{\rho_1(x)},$$

for all $f \in C_{\rho_1}$ and $s \in \mathbb{R}$.

Lemma 1. *Suppose that for positive linear operators $A_n : C_{\rho_1} \longrightarrow B_{\rho_2}$ the sequence $\|A_n\|_{C_{\rho_1}\to B_{\rho_1}}$ of operator norms is uniformly bounded, and*

$$\lim_{x\to\infty} \frac{\rho_1(x)}{\rho_2(x)} = 0 \tag{1}$$

and $\lim_{n\to\infty} \psi_n(s) = 0$ for any s. Then

$$\lim_{n\to\infty} \|A_n\|_{C_{\rho_1}\longrightarrow B_{\rho_2}} = 0.$$

Proof. Since $\lim_{x\to\infty} \frac{\rho_1(x)}{\rho_2(x)} = 0$, there exists a number s_0 for $\varepsilon > 0$ such that $\frac{\rho_1(x)}{\rho_2(x)} \leq \varepsilon$ for all $|x| > s_0$. Since ρ_1 and ρ_2 are continuous and strictly positive, the function $\frac{\rho_1}{\rho_2}$ is also continuous and bounded for $|x| \leq s_0$. Then there exists a number $c_1 > 0$ such that $\frac{\rho_1(x)}{\rho_2(x)} \leq c_1$ for all $|x| \leq s_0$. On the other side there exists, from the hypothesis, a number $c_2 > 0$ such that $\|A_n\|_{C_{\rho_1}\longrightarrow B_{\rho_2}} \leq c_2$ for all $n \in \mathbb{N}$. Hence we have the inequality

$$\begin{aligned}
\|A_n\|_{C_{\rho_1}\longrightarrow B_{\rho_2}} &= \sup_{\|f\|_{\rho_1}=1} \left\{ \sup_{x\in\mathbb{R}} \frac{|A_n(f,x)|}{\rho_2(x)} \right\} \\
&\leq \sup_{\|f\|_{\rho_1}=1} \left\{ \sup_{|x|>s_0} \frac{|A_n(f,x)|}{\rho_1(x)} \frac{\rho_1(x)}{\rho_2(x)} \right\} + \sup_{\|f\|_{\rho_1}=1} \left\{ \sup_{|x|\leq s_0} \frac{|A_n(f,x)|}{\rho_1(x)} \frac{\rho_1(x)}{\rho_2(x)} \right\} \\
&\leq \varepsilon \|A_n\|_{C_{\rho_1}\longrightarrow B_{\rho_1}} + c_1 \cdot \psi_n(s_0) \\
&\leq \varepsilon c_2 + c_1 \cdot \psi_n(s_0).
\end{aligned}$$

From this and the hypothesis we obtain

$$\lim_{n\to\infty} \|A_n\|_{C_{\rho_1}\longrightarrow B_{\rho_2}} = 0$$

and the proof is complete. □

Theorem 1. *Let the weight functions ρ_1 and ρ_2 be as in Lemma 1 and let the sequence $\|L_n\|_{C_{\rho_1}\to B_{\rho_1}}$ of operator norms be uniformly bounded. Here $L_n : C_{\rho_1} \longrightarrow B_{\rho_2}$ are positive linear operators. If the equality*

$$\lim_{n\to\infty} |L_n(f,x) - f(x)| = 0$$

holds for all s_0 with $|x| \leq s_0$, then

$$\lim_{n\to\infty} \|L_n(f,x) - f(x)\|_{\rho_2} = 0$$

for all $f \in C_{\rho_1}$.

Proof. Let E be the identity operator on C_{ρ_1}, and replace the operators A_n in Lemma 1 by $L_n - E$. Since $\rho_1(x) \geq 1$ for all x, we obtain the inequality

$$\psi_n(s_0) \leq \sup_{\|f\|_{\rho_1}=1} \left\{ \sup_{|x|\leq s_0} |L_n(f,x) - f(x)| \right\}.$$

By hypotheses it follows that

$$\lim_{n\to\infty} \psi_n(s_0) = 0.$$

Hence from Lemma 1 it follows $\lim_{n\to\infty} \|L_n - E\|_{C_{\rho_1}\longrightarrow B_{\rho_2}} = 0$, and so

$$\|L_n(f,x) - f(x)\|_{\rho_2} \leq \|L_n - E\|_{C_{\rho_1}\longrightarrow B_{\rho_2}} \|f\|_{\rho_1}.$$

From this we obtain the required result. □

Remark. Let (A_n), $A_n : C_{\rho_1} \longrightarrow B_{\rho_2}$, be a sequence of positive linear operators for all $n \in \mathbb{N}$. Suppose that there exists $M > 0$ such that for all $x \in \mathbb{R}$ we have $\rho_1(x) \leq M\rho_2(x)$. If

$$\lim_{n\to\infty} \|A_n(\rho_1, x) - \rho_1(x)\|_{\rho_2} = 0,$$

then the sequence $(A_n)_{n\in\mathbb{N}}$ is uniformly bounded.

Let φ_1 and φ_2 be two continuous functions, monotonically increasing on the real axis, such that $\lim_{x\to\mp\infty} \varphi_1(x) = \lim_{x\to\mp\infty} \varphi_2(x) = \mp\infty$ and $\rho_k(x) = 1 + \varphi_k^2(x)$, $k = 1,2$.

Theorem 2. *If the positive linear operators sequence*

$$A_n : C_{\rho_1} \longrightarrow B_{\rho_2}$$

satisfies the following three conditions

$$\lim_{n\to\infty} \|A_n(\varphi_1^\nu, x) - \varphi_1^\nu(x)\|_{\rho_2} = 0, \qquad \nu = 0,1,2, \tag{2}$$

and the condition expressed in equation (1). *Then*

$$\lim_{n\to\infty} \|A_n(f,x) - f(x)\|_{\rho_2} = 0$$

for all $f \in C_{\rho_1}$.

Proof. To prove the theorem, it is sufficient to show that the conditions of Theorem 1 should be satisfied, i.e, the sequence of operator norms of A_n is uniformly bounded and $\lim_{n\to\infty} |A_n(f,x) - f(x)| = 0$ for $|x| \leq s_0$. Let us show that the sequence of operator norms of A_n is uniformly bounded. From the hypotheses

$$\lim_{n\to\infty} \|A_n(1,x) - 1\|_{\rho_2} = 0$$

and

$$\lim_{n\to\infty} \|A_n(\varphi_1^2(x), x) - \varphi_1^2(x)\|_{\rho_2} = 0.$$

It follows that

$$\begin{aligned}\lim_{n\to\infty}\|A_n(\rho_1,x)-\rho_1\|_{\rho_2}\leq\lim_{n\to\infty}\|A_n(1,x)-1\|_{\rho_2}\\+\lim_{n\to\infty}\|A_n(\varphi_1^2(x),x)-\varphi_1^2(x)\|_{\rho_2}=0.\end{aligned}$$

That means the sequence of operator norms are uniformly bounded from the Remark above.

Now let us examine the difference $|A_n(f,x)-f(x)|$:

$$\begin{aligned}|A_n(f,x)-f(x)|&\leq A_n(|f(t)-f(x)|,x)+|f(x)||A_n(1,x)-1|\\&=I_n'(x)+I_n''(x).\end{aligned}$$

Firstly, we investigate the limit of $I_n''(x)$ for $n\to\infty$.

Since $f(x)$ is a continuous function, it is bounded on the interval $|x|\leq s_0$ for any s_0. Now set $M_1:=\max_{|x|\leq s_0}|f(x)|$. Note that from the hypothesis we have

$$\lim_{n\to\infty}\|A_n(1,x)-1\|_{\rho_2}=\lim_{n\to\infty}\sup_{x\in\mathbb{R}}\frac{|A_n(1,x)-1|}{\rho_2(x)}=0.$$

Hence, for a zero sequence, ε_n

$$\sup_{x\in\mathbb{R}}\frac{|A_n(1,x)-1|}{\rho_2(x)}=\varepsilon_n.$$

From this we have

$$|A_n(1,x)-1|\leq\varepsilon_n\rho_2(x)$$

for all $x\in\mathbb{R}$, and so

$$\lim_{n\to\infty}|A_n(1,x)-1|=0.$$

Therefore this implies

$$\lim_{n\to\infty}I_n''(x)=\lim_{n\to\infty}|f(x)||A_n(1,x)-1|\leq\lim_{n\to\infty}M_1|A_n(1,x)-1|=0$$

for all $|x|\leq s_0$.

Let us obtain some inequalities that can be used to find the limit of I_n' as n tends to infinity. It is easy to see that the inequality

$$\begin{aligned}|f(t)-f(x)|&\leq 2M_f\rho_1(t)\rho_1(x)\\&\leq 4M_f\rho_1(x)[1+(\varphi_1(t)-\varphi_1(x))^2+\varphi_1^2(x)]\end{aligned}\tag{$*$}$$

holds. Setting

$$\Delta_\rho(\varphi_1,x):=\min\{\varphi_1(x+\delta)-\varphi_1(x);\ \varphi_1(x)-\varphi_1(x-\delta)\},\tag{3}$$

we obtain

$$|\varphi_1(t)-\varphi_1(x)|>\min\{\varphi_1(x+\delta)-\varphi_1(x);\varphi_1(x)-\varphi_1(x-\delta)\},$$

and therefore

$$\frac{1}{|\varphi_1(t)-\varphi_1(x)|}<\frac{1}{\Delta_\rho(\varphi_1,x)}.$$

$\rho_1(x) \geq 1$ and the inequality (*) implies

$$|f(t) - f(x)| < 4M_f\rho_1(x)(\varphi_1(t) - \varphi_1(x))^2\left[\frac{1}{\Delta_\rho^2(\varphi_1,x)} + 1 + \frac{\varphi_1(x)^2}{\Delta_\rho^2(\varphi_1,x)}\right]$$
$$\leq 4M_f\rho_1(x)^2(\varphi_1(t) - \varphi_1(x))^2\left[\frac{1}{\Delta_\rho^2(\varphi_1,x)} + 1\right].$$

On the other hand, from the continuity of the function $f(x)$, there exists a number $\varepsilon > 0$ such that

$$|f(t) - f(x)| < \varepsilon$$

for all t and x for which $|t - x| < \delta$. If we set

$$K_{\rho_1}(x) := 4M_f\rho_1(x)^2\left[1 + \frac{1}{\Delta_\rho^2(\varphi_1,x)}\right],$$

then we obtain

$$|f(t) - f(x)| < \varepsilon + K_{\rho_1}(x)(\varphi_1(t) - \varphi_1(x))^2 \tag{4}$$

for all $t \in \mathbb{R}$ and x for which $|x| \leq s_0$. Since the function φ_1 is monotonically increasing, then $\Delta_\rho(\varphi_1,x) \neq 0$, and therefore $K_{\rho_1}(x)$ is a continuous function. Now suppose

$$M_2 := \max_{|x|\leq s_0} K_{\rho_1}(x).$$

The monotonicity of the operators A_n and eq. (4) yield the inequality

$$A_n(|f(t) - f(x)|,x) \leq \varepsilon[A_n(1,x) - 1] + \varepsilon + M_2A_n((\varphi_1(t) - \varphi_1(x))^2,x).$$

We obtain

$$|A_n(1,x) - 1| < \varepsilon_n\rho_2(x),$$
$$|A_n(\varphi,x) - \varphi(x)| < \varepsilon_n\rho_2(x),$$
$$|A_n(\varphi^2,x) - \varphi^2(x)| < \varepsilon_n\rho_2(x),$$

from the hypothesis

$$\lim_{n\to\infty} \|A_n(\varphi^\nu(x),x) - \varphi^\nu(x)\|_{\rho_2} = 0, \quad \nu = 0,1,2.$$

We can write from these inequalities that

$$A_n((\varphi_1(t) - \varphi_1(x))^2,x) < \varepsilon_n\rho_2(x)(1 + \varphi_1(x))^2.$$

The boundness of $\rho_k(x)$ $(k = 1,2)$ for $|x| \leq s_0$ yields

$$\lim_{n\to\infty} A_n((\varphi_1(t) - \varphi_1(x))^2,x) < \lim_{n\to\infty} \varepsilon_n\rho_1^2(x)\rho_2(x) = 0.$$

Hence

$$\lim_{n\to\infty} I_n'(x) \leq \lim_{n\to\infty}[\varepsilon(A_n(1,x) - 1) + \varepsilon + M_2A_n((\varphi_1(t) - \varphi_1(x))^2,x)]$$
$$< \lim_{n\to\infty}[\varepsilon \cdot \varepsilon_n \cdot \rho_2(x) + M_2A_n((\varphi_1(t) - \varphi_1(x))^2,x)] = 0$$

for all $|x| \leq s_0$. This proves the theorem. □

Note. Lemma 1, Theorems 1 and 2 have been obtained by Gadjiev, cf. [5], for $\rho_1 = \rho_2$. The results in ref. [5] cannot be deduced from our theorems from the condition (1).

Acknowledgements

The author would like to thank Prof. A D Gadjiev for several helpful discussions and comments.

References

[1] Altomare F and Campiti M, *Korovkin-type approximation theory and its applications* (Berlin, New York: Walter de Gruyter) (1994)

[2] Baskakov V A, *On a sequence of linear positive operators* In Research in modern constructive functions theory (Moscow) (1961)

[3] Coşkun T, Some properties of linear positive operators on the spaces of weight functions, *Commun. Fac. Sci. Univ. Ank. Series. A1* **47** (1998) 175–181

[4] Gadjiev A D, The convergence problem for a sequence of positive linear operators on unbounded sets and theorems analogous to that of P P Korovkin. English translated, *Sov. Math. Dokl.* **15** (1974) 5

[5] Gadjiev A D, On P P Korovkin type theorems, *Mathem. Zametki* **20** (1976) 5 (in Russian)

[6] Korovkin P P, *Linear operators and approximation theory* (Delhi) (1960)

Proc. Indian Acad. Sci. (Math. Sci.), Vol. 110, No. 4, November 2000, pp. 363–377.
© Printed in India

Hyperfinite representation of distributions

J SOUSA PINTO* and R F HOSKINS†

*Departamento de Matemática, Universidade de Aveiro, Aveiro, Portugal
†Department of Mathematical Sciences, De Montfort University, Leicester, UK

MS received 28 February 2000; revised 30 May 2000

Abstract. Hyperfinite representation of distributions is studied following the method introduced by Kinoshita [2, 3], although we use a different approach much in the vein of [4]. Products and Fourier transforms of representatives of distributions are also analysed.

Keywords. Generalized functions; Schwartz distributions; nonstandard analysis; hyperfinite representations.

1. Notation and preliminary results

A nonstandard treatment of the theory of distributions in terms of a hyperfinite representation has been presented in papers [2, 3] by Kinoshita. A further exploitation of this treatment in an N-dimensional context has been given by Grenier [1]. In the present paper we offer a different approach to the hyperfinite representation, based on the nonstandard theory of distributions developed in [4]. Some basic acquaintance with nonstandard analysis (NSA) is assumed. For the most part little more is needed than what is contained in the description in [4] of an elementary ultrapower model of the hyperreals. For a more detailed study of the fundamentals of NSA see, for example, Luxemburg [6] or Lindstrøm [5].

Let κ be any given infinite hypernatural number which, without any loss of generality, will be supposed to be even; then define $\varepsilon = \kappa^{-1} \approx 0$. Hence,

$$\Pi \equiv \Pi_\kappa = \left\{-\frac{\kappa}{2}, -\frac{\kappa}{2} + \varepsilon, \ldots, 0, \ldots, \frac{\kappa}{2} - \varepsilon\right\}$$

$$= \left\{\left(-\frac{\kappa}{2} + j - 1\right)\varepsilon : j = 1, 2, \ldots, \kappa^2\right\} \subset {}^*\mathbb{R}$$

is an (internal) hyperfinite set of hyperreal numbers with internal cardinality κ^2. Π will be referred to as the (unbounded) *hyperfinite line*. Given a standard point $r \in \mathbb{R}$, define the Π-monad of r by

$$\mathrm{mon}_\Pi(r) = \mathrm{st}_\Pi^{-1}(r) = \mathrm{mon}(r) \cap \Pi,$$

where mon denotes the usual monad of a standard number in ${}^*\mathbb{R}$. Then the set $\Pi_b = \cup_{r\in\mathbb{R}} \mathrm{mon}_\Pi(r) = \mathrm{st}_\Pi^{-1}(\mathbb{R}) \subset \Pi$ is the *nearstandard hyperfinite line* and $\Pi_\infty = \Pi \backslash \Pi_b$ is the set of *remote points* of the hyperfinite line. For every subset $A \subset \mathbb{R}$ define ${}^*A_\Pi = {}^*A \cap \Pi$ and $\mathrm{ns}_\Pi({}^*A) = {}^*A \cap \Pi_b = \cup_{a\in A} \mathrm{mon}_\Pi(a)$. The notation throughout will be the usual in the field.

Now consider the basic set of internal functions

$$^{\Pi}\mathbb{F} = \{F : \Pi \to {}^*\mathbb{C} : F \text{ is internal}\}$$

and suppose, if necessary, that each $F \in {}^{\Pi}\mathbb{F}$ is periodically extended to the infinite grid $\varepsilon \cdot {}^*\mathbb{Z}$. Defining addition and scalar multiplication componentwise, ${}^{\Pi}\mathbb{F}$ is a ${}^*\mathbb{C}$-linear space of hyperfinite dimension κ^2. Moreover, defining also componentwise the product of two functions, ${}^{\Pi}\mathbb{F}$ is in fact an algebra. The operators $\mathbf{D}_+, \mathbf{D}_- : {}^{\Pi}\mathbb{F} \to {}^{\Pi}\mathbb{F}$ defined, for every function F and $x \in \Pi$, by

$$\mathbf{D}_+F(x) = \varepsilon^{-1}[F(x+\varepsilon) - F(x)] \text{ and } \mathbf{D}_-F(x) = \varepsilon^{-1}[F(x) - F(x-\varepsilon)]$$

are called, respectively, the forward and the backward Π-difference operators (of first order). Iterating $\mathbf{D}_+$ (or $\mathbf{D}_-$) we obtain higher order Π-difference operators: for every (finite or infinite) $n \in {}^*\mathbb{N}_0$

$$\mathbf{D}_+^n F(x) = \mathbf{D}_+(\mathbf{D}_+^{n-1}F(x)), \quad x \in \Pi$$

and similarly for $\mathbf{D}_-^n$. It is easily seen that for any two functions $F, G \in {}^{\Pi}\mathbb{F}$ we have (both for $\mathbf{D}_+$ and $\mathbf{D}_-$),

$$\mathbf{D}(F+G) = \mathbf{D}F + \mathbf{D}G \text{ and } \mathbf{D}(F \cdot G) = (\mathbf{D}F)G + F(\mathbf{D}G) \pm \varepsilon(\mathbf{D}F)(\mathbf{D}G),$$

where we take $+\varepsilon$ or $-\varepsilon$ according as we use $\mathbf{D}_+$ or $\mathbf{D}_-$, respectively.

For every $\alpha, x \in \Pi$ define the Π-intervals (containing only points in Π) $J_\alpha^+(x)$ and $J_\alpha^-(x)$ as follows:

$$J_\alpha^+(x) = \begin{cases} (x,\alpha]_\Pi & \text{if } x < \alpha \\ [x,\alpha)_\Pi & \text{if } x > \alpha \end{cases}$$
$$J_\alpha^-(x) = \begin{cases} [\alpha,x)_\Pi & \text{if } x < \alpha \\ (\alpha,x]_\Pi & \text{if } x > \alpha \end{cases}$$

while for $x = \alpha$ we have $J_\alpha^+(x) = \emptyset = J_\alpha^-(x)$. For any $F \in {}^{\Pi}\mathbb{F}$ define the functions $\mathbf{S}_+F$ and $\mathbf{S}_-F$ to be the forward and backward Π-sums of F which are zero at the origin and which, for every $x \in \Pi\backslash\{0\}$ are defined by

$$\mathbf{S}_+F(x) = \sum_{t \in J_0^+(x)} \varepsilon F(t) \text{ and } \mathbf{S}_-F(x) = \sum_{t \in J_0^-(x)} \varepsilon F(t).$$

The Π-*sum* operators $\mathbf{S}_+$ and $\mathbf{S}_-$ both transform ${}^{\Pi}\mathbb{F}$ into ${}^{\Pi}\mathbb{F}$. Moreover, for every $F \in {}^{\Pi}\mathbb{F}$, we have

$$\mathbf{D}_+\mathbf{S}_+F = F \quad \text{and} \quad \mathbf{D}_-\mathbf{S}_-F = F$$

that is, $\mathbf{S}_+$ and $\mathbf{S}_-$ are left inverses for $\mathbf{D}_+$ and $\mathbf{D}_-$, respectively.

1.1 SΠ-*continuous functions*

Given a (standard) function $f : A \to \mathbb{C}$ defined on a subset A of $\mathbb{R}$ we always consider its extension to the whole of $\mathbb{R}$, denoted again by f, by setting $f(x) = 0$ on $A^c \equiv \mathbb{R}\backslash A$. For any such function consider the nonstandard extension ${}^*f$ and then define ${}^*f_\Pi$ to be the restriction of ${}^*f$ to Π (periodically extended to $\varepsilon \cdot {}^*\mathbb{Z}$). Hence, for every standard function f, we clearly have ${}^*f_\Pi \in {}^{\Pi}\mathbb{F}$.

DEFINITION 1.1

An internal function $F \in {}^{\Pi}\mathbb{F}$ is said to be SΠ-continuous on a nonempty subset Ω of Π if and only if

$$\forall_{x,y}[x, y \in \Omega \quad \text{and} \quad x \approx y \Rightarrow F(x) \approx F(y)].$$

From the nonstandard characterization of (standard) continuity and uniform continuity there follows

Theorem 1.2. *If $f : \mathbb{R} \to \mathbb{C}$ is a (standard) function which is continuous at a point $r \in \mathbb{R}$ then ${}^{*}f_{\Pi} : \Pi \to {}^{*}\mathbb{C}$ is* SΠ*-continuous on* $\mathrm{mon}_{\Pi}(r)$*. If f is continuous on the set $A \subset \mathbb{R}$ then ${}^{*}f_{\Pi}$ is* SΠ*-continuous on* $\mathrm{ns}_{\Pi}({}^{*}A)$*. Moreover, if f is uniformly continuous on A, then ${}^{*}f_{\Pi}$ is* SΠ*-continuous on ${}^{*}A_{\Pi}$.*

The converse does not necessarily hold. The internal function ${}^{*}f_{\Pi}$ may have infinitesimal variation over the Π-monad of a (standard) point, but this fact does not ensure that the variation is kept at an infinitesimal level over the entire monad of the same point. Consider, for example, the (standard) Dirichlet function

$$\mathbf{d}(x) = \begin{cases} 1 & \text{if } x \in [0,1]\backslash\mathbb{Q} \\ 0 & \text{otherwise.} \end{cases}$$

Since Π contains only hyperrational points, ${}^{*}\mathbf{d}_{\Pi}$ is zero for all $x \in \Pi$ and it follows that ${}^{*}\mathbf{d}_{\Pi}$ is SΠ-continuous on Π_b while ${}^{*}\mathbf{d}$ is not S-continuous anywhere on ${}^{*}[0,1] \subset {}^{*}\mathbb{R}$.

Consider an internal function $F \in {}^{\Pi}\mathbb{F}$ such that

(a) $F(x)$ is finite on Π_b, and
(b) F is SΠ-continuous on Π_b.

Then it make sense to define the (standard) function $\mathrm{st}F : \mathbb{R} \to \mathbb{C}$ by setting for every $t \in \mathbb{R}$

$$\mathrm{st}F(t) = [\mathrm{st} \circ F](x), \quad \text{for any } x \in \mathrm{mon}_{\Pi}(t).$$

If ι is any *choice function* picking up one and only one point from each set to which it is applied then we may write $\mathrm{st}F = \mathrm{st} \circ F \circ \iota \circ \mathrm{st}_{\Pi}^{-1}$.

Denote by $\mathbf{SC}_{\Pi} \equiv \mathbf{SC}_{\Pi}(\mathbb{R})$ the set of all functions in ${}^{\Pi}\mathbb{F}$ which are finite and SΠ-continuous on Π_b.

As the above example concerning the Dirichlet function shows we cannot expect in the general case to recover the original function $f : \mathbb{R} \to \mathbb{C}$ from its Π-extension. However, it is not difficult to see that

Theorem 1.3. *If f is a continuous function on a subset A of $\mathbb{R}$ then we have* $\mathrm{st}({}^{*}f_{\Pi}) = f$ *on A.*

and, more generally,

Theorem 1.4. *If f is a function which is k times continuously differentiable on a subset A of $\mathbb{R}$ then* $\mathrm{st}\,\mathbf{D}^{j}_{+}({}^{*}f_{\Pi}) = f^{(j)}$*, $j = 0, 1, 2, \ldots, k$, hold on A. (The same holds if we consider $\mathbf{D}^{j}_{-}$ instead of $\mathbf{D}^{j}_{+}$.)*

2. SΠ-distributions

Given any $F \in \mathbf{SC}_\Pi$, the function $\mathbf{S}_+F$ (or $\mathbf{S}_-F$) is again in $\mathbf{SC}_\Pi$. In fact, for any $x \in \Pi_b$, considering $\mathbf{S}_+F$ for example, we have

$$|\mathbf{S}_+F(x)| \leq \sum_{t \in J_0^+(x)} \varepsilon |F(t)| \leq \left\{ \max_{t \in J_0^+(x)} |F(t)| \right\} |x|$$

and thus $\mathbf{S}_+F$ is finite on Π_b. Also, for any $x, y \in \Pi$,

$$|\mathbf{S}_+F(y) - \mathbf{S}_+F(x)| \leq \sum_{t \in J^+_{\{x,y\}}} \varepsilon\, |F(t)| \leq |y - x| \cdot \max_{t \in J^+_{\{x,y\}}} |F(t)|,$$

where $J^+_{\{x,y\}} \equiv J^+_{\min\{x,y\}}(\max\{x,y\})$. Hence, if $x \approx y$ then $\mathbf{S}_+F(x) \approx \mathbf{S}_+F(y)$ and therefore $\mathbf{S}_+F \in \mathbf{SC}_\Pi$, as asserted.

The same result is not generally true for Π-differences. If $F \in \mathbf{SC}_\Pi$ then the most we can say about the function $\mathbf{D}_+F$ (or $\mathbf{D}_-F$), in principle, is that it belongs to ${}^\Pi\mathbb{F}$.

2.1 *The* ${}^*\mathbb{C}_b$*-module* ${}^\Pi\mathbb{D}_\infty$

For any F in $\mathbf{SC}_\Pi$, $\operatorname{st} F$ is a (standard) continuous function on $\mathbb{R}$ which therefore defines a (regular) distribution in $\mathcal{D}'$. Denoting by ν_F either the function $\operatorname{st} F$ or the distribution it generates as the context demands, we have

$$\langle \nu_F, \varphi \rangle = \int_{\mathbb{K}} \nu_F(t)\varphi(t)\mathrm{d}t = \int_{\mathrm{st}_\Pi^{-1}(\mathbb{K})} (\mathrm{st} \circ F)\varphi_\Pi \mathrm{d}\Lambda_L, \tag{1}$$

where $\mathrm{st} \circ F$ and $\varphi_\Pi = \mathrm{st} \circ {}^*\varphi_\Pi$ are (external) functions defined on Π, $\mathbb{K}$ is a compact of $\mathbb{R}$ containing the support of φ and Λ_L denotes the counting Loeb measure on Π. Since $F \cdot {}^*\varphi_\Pi$ is an SΠ-lifting for the external function $(\mathrm{st} \circ F)\varphi_\Pi$ we may replace the Loeb integral in (1) by a proper Π-sum to obtain

$$\langle \nu_F, \varphi \rangle = \mathrm{st}\left(\sum_{x \in {}^*\mathbb{K}_\Pi} \varepsilon\, F(x)\, {}^*\varphi_\Pi(x) \right).$$

It is easy to see that $\varphi \rightsquigarrow {}^*\varphi_\Pi$ is a linear and continuous map and therefore every internal function $F \in \mathbf{SC}_\Pi$ generates in this way a regular distribution. Since the map $f \rightsquigarrow {}^*f_\Pi$ embeds $\mathcal{C} \equiv \mathcal{C}(\mathbb{R})$ into $\mathbf{SC}_\Pi$ and the distribution generated by f coincides with $\nu_{{}^*f_\Pi}$, the map

$$\mathrm{st}_{\mathcal{D}} : \mathbf{SC}_\Pi \to \mathcal{D}'$$

defined by $\mathrm{st}_{\mathcal{D}}(F) = \nu_F$, establishes an onto correspondence between $\mathbf{SC}_\Pi$ and the subspace of $\mathcal{D}'$ comprising all regular distributions generated by continuous functions on $\mathbb{R}$.

Now, if $F \in \mathbf{SC}_\Pi$ and $\varphi \in \mathcal{D}$ is a function with support in the compact $\mathbb{K} \sqsubset \mathbb{R}$ then, taking Theorem 1.4 into account, we get

$$\begin{aligned}
\sum_{x \in \Pi} \varepsilon\, \mathbf{D}_+F(x)^*\varphi_\Pi(x) &= \sum_{x \in \Pi} [F(x+\varepsilon) - F(x)]^*\varphi_\Pi(x) \\
&= \sum_{x \in \Pi} \varepsilon F(x)(-\mathbf{D}_-{}^*\varphi_\Pi(x)) \approx \int_{\mathrm{st}_\Pi^{-1}(\mathbb{K})} (\mathrm{st} \circ F)(-\varphi')_\Pi \mathrm{d}\Lambda_L \\
&= \langle \nu_F, -\varphi' \rangle = \langle \mathbf{D}\nu_F, \varphi \rangle,
\end{aligned}$$

where $\mathbf{D}\nu_F$ is the (standard) distributional derivative of ν_F. Let $\mathbf{D}_+(\mathbf{SC}_\Pi)$ be the set of first order Π-differences of all functions in $\mathbf{SC}_\Pi$. Since for every $F \in \mathbf{SC}_\Pi$ we have that $F = \mathbf{D}_+(\mathbf{S}_+F)$ where $\mathbf{S}_+F \in \mathbf{SC}_\Pi$ then $\mathbf{SC}_\Pi \subset \mathbf{D}_+(\mathbf{SC}_\Pi)$. Then the $\mathrm{st}_{\mathcal{D}}$-mapping may be extended onto $\mathbf{D}_+(\mathbf{SC}_\Pi)$, by setting

$$\mathrm{st}_{\mathcal{D}}(\mathbf{D}_+F) = \mathbf{D}\nu_F = \mathbf{D}(\mathrm{st}_{\mathcal{D}}(F)).$$

The same idea may be generalized to Π-differences of any *finite* order of a function in $\mathbf{SC}_\Pi$. Hence, if $F \in \mathbf{SC}_\Pi$ and $\varphi \in \mathcal{D}$, we obtain, for every $j \in \mathbb{N}_0$,

$$\begin{aligned}\sum_{x\in\Pi} \varepsilon\, \mathbf{D}^j_+F(x)^\star\varphi_\Pi(x) &= \sum_{x\in\Pi} \varepsilon\, F(x)[(-1)^j\mathbf{D}^j_-{}^\star\varphi_\Pi(x)] \\ &\approx \int_{\mathrm{st}_\Pi^{-1}(\mathbb{K})} (\mathrm{st}\circ F)(-\varphi^{(j)})_\Pi\, \mathrm{d}\Lambda_L \\ &= \langle \nu_F, (-1)^j\varphi^{(j)}\rangle = \langle \mathbf{D}^j(\nu_F), \varphi\rangle,\end{aligned}$$

that is, $\mathrm{st}_{\mathcal{D}}(\mathbf{D}^j_+F) = \mathbf{D}^j(\mathrm{st}_{\mathcal{D}}(F))$.

Denoting by $\mathbf{D}^j_+(\mathbf{SC}_\Pi)$, for every $j \in \mathbb{N}_0$, the set of $\mathbf{D}^j_+$-differences of all functions in $\mathbf{SC}_\Pi$, then we have the inclusion $\mathbf{D}^j_+(\mathbf{SC}_\Pi) \subset \mathbf{D}^{j+1}_+(\mathbf{SC}_\Pi)$, and therefore

$${}^\Pi\mathbb{D}_\infty \equiv {}^\Pi\mathbb{D}_\infty(\mathbb{R}) = \bigcup_{j=0}^{\infty} \mathbf{D}^j_+(\mathbf{SC}_\Pi)$$

is the (external) set of all finite-order Π-differences of all functions in $\mathbf{SC}_\Pi$. Since for every $F \in \mathbf{SC}_\Pi$ the translate $\tau_\varepsilon F$ is also in $\mathbf{SC}_\Pi$ and, moreover, $\mathbf{D}_- = \mathbf{D}_+ \circ \tau_\varepsilon$ then ${}^\Pi\mathbb{D}_\infty$ may be obtained using indifferently either $\mathbf{D}_+$ or $\mathbf{D}_-$. Hence we may also write, more generally,

$${}^\Pi\mathbb{D}_\infty \equiv {}^\Pi\mathbb{D}_\infty(\mathbb{R}) = \bigcup_{j=0}^{\infty}\bigcup_{k=0}^{\infty} \mathbf{D}^j_+\mathbf{D}^k_-(\mathbf{SC}_\Pi).$$

We may now extend the map $\mathrm{st}_{\mathcal{D}}$ to the whole of ${}^\Pi\mathbb{D}_\infty$ as follows: for every $\Phi \in {}^\Pi\mathbb{D}_\infty$ there exist $F \in \mathbf{SC}_\Pi$ and $j \in \mathbb{N}_0$ so that $\Phi = \mathbf{D}^j_+F$. Hence, $\mathrm{st}_{\mathcal{D}}(\Phi) = \mathbf{D}^j\nu_F \in \mathcal{D}'$. Note that $\mathrm{st}_{\mathcal{D}}(\Phi)$ does not depend upon the representation of Φ as a finite order Π-difference of a function in $\mathbf{SC}_\Pi$. In fact, suppose we also have $\Phi = \mathbf{D}^m_+G$ with $G \in \mathbf{SC}_\Pi$ and $m \in \mathbb{N}_0$ (where, without any loss of generality we may assume $m \geq j$). Then from the equation $\mathbf{D}^j_+F = \mathbf{D}^m_+G$ it follows that $\mathbf{S}^{m-j}_+F + P_m = G$, where P_m is a polynomial of degree $< m$ (and coefficients in ${}^*\mathbb{C}$). Thus, for any $\varphi \in \mathcal{D}$, we get

$$\begin{aligned}\langle \mathbf{D}^m\nu_G, \varphi\rangle &= \langle \nu_G, (-1)^m\varphi^{(m)}\rangle \\ &\approx \sum_{x\in\Pi} \varepsilon\, G(x)[(-1)^m\mathbf{D}^{m\star}_-\varphi_\Pi(x)] \\ &= \sum_{x\in\Pi} \varepsilon[\mathbf{S}^{m-j}_+F(x) + P_m(x)][(-1)^m\mathbf{D}^{m\star}_-\varphi_\Pi(x)] \\ &= \sum_{x\in\Pi} \varepsilon\, F(x)(-1)^j\mathbf{D}^j_-{}^\star\varphi_\Pi(x) + \sum_{x\in\Pi} \varepsilon\, \mathbf{D}^m_+P_m(x)^\star\varphi_\Pi(x) \\ &\approx \langle \nu_F, (-1)^j\varphi^{(j)}\rangle = \langle \mathbf{D}^j\nu_F, \varphi\rangle\end{aligned}$$

and therefore $\mathbf{D}^m\nu_G = \mathbf{D}^j\nu_F$ which proves the assertion made.

The *$\mathcal{D}$-standard part map* $\mathrm{st}_{\mathcal{D}} : {}^\Pi\mathbb{D}_\infty \to \mathcal{D}'$ is clearly linear; its kernel, $\mathcal{K}_\infty \equiv \mathcal{K}_\infty(\mathrm{st}_{\mathcal{D}})$, comprises all internal functions in ${}^\Pi\mathbb{D}_\infty$ which generate the null distribution. These are all the functions which are finite order derivatives of infinitesimal functions in $\mathbf{SC}_\Pi$. The factor space ${}^{\Pi}\mathcal{C}_\infty \equiv {}^{\Pi}\mathbb{D}_\infty/\mathcal{K}_\infty$ is a $\mathbb{C}$-vector space which may be shown to be isomorphic to $\mathcal{C}_\infty$, the space of all finite order Schwartz distributions.

DEFINITION 2.1

The internal functions in ${}^{\Pi}\mathbb{D}_\infty \subset {}^{\Pi}\mathbb{F}$ will be called finite order Π-predistributions and the classes $[\Phi] \in {}^{\Pi}\mathcal{C}_\infty$, with $\Phi \in {}^{\Pi}\mathbb{D}_\infty$, will be called *finite order* $S\Pi$*-distributions.*

The Π-predistributions are internal functions in ${}^{\Pi}\mathbb{F}$ which do not grow too fast on Π_b according to the following result:

Theorem 2.2. *For every internal function* $\Phi \in {}^{\Pi}\mathbb{D}_\infty$ *there exists a finite nonnegative integer* $m \equiv m_\Phi$ *such that for every compact* $\mathbb{K}$ *of* $\mathbb{R}$

$$|\Phi(x)| \leq \mathbf{C}_{\mathbb{K},\Phi} \cdot \kappa^m, \quad \text{on } {}^*\mathbb{K}_\Pi \tag{2}$$

where $\mathbf{C}_{\mathbb{K},\Phi}$ *is a finite positive constant (depending on* $\mathbb{K}$ *and* Φ*).*

Proof. The inequality (2) clearly holds for every $\Phi \in \mathbf{SC}_\Pi$ with $m = 0$. Now, if we have $\Phi = \mathbf{D}_+F$ with $F \in \mathbf{SC}_\Pi$, then

$$\Phi(x) = \mathbf{D}_+F(x) = \kappa\,[F(x+\varepsilon) - F(x)]$$

and therefore, for every compact $\mathbb{K} \sqsubset \mathbb{R}$, we obtain

$$\max_{x\in{}^*\mathbb{K}_\Pi} |\Phi(x)| \leq 2\left\{\max_{x\in{}^*\mathbb{K}_\Pi} |F(x)|\right\} \cdot \kappa.$$

Hence the inequality holds with $C_{\mathbb{K},\Phi} = 2\max_{x\in{}^*\mathbb{K}_\Pi} |F(x)| \in {}^*\mathbb{R}_b$ and $m = 1$.

Suppose now that the inequality holds for all internal functions of the form $\mathbf{D}^j_+F$ with $F \in \mathbf{SC}_\Pi$. If $\Phi = \mathbf{D}^{j+1}_+F$ with $F \in \mathbf{SC}_\Pi$ then we obtain,

$$\max_{x\in{}^*\mathbb{K}_\Pi} |\Phi(x)| \leq 2\left\{\max_{x\in{}^*\mathbb{K}_\Pi} |\mathbf{D}^j_+F(x)|\right\} \cdot \kappa \leq C^{[j+1]}_{\mathbb{K},\Phi} \cdot \kappa^{j+1},$$

where $C^{[j+1]}_{\mathbb{K},\Phi}$ is, for every fixed $j \in \mathbb{N}_0$, a positive bounded constant. Therefore the result follows by finite induction. Note that there are functions $F \in \mathbf{SC}_\Pi$ such that $\mathbf{D}_+F \in \mathbf{SC}_\Pi$; then, for a general function of the form $\Phi = \mathbf{D}^j_+F$ with $F \in \mathbf{SC}_\Pi$, equation (2) may be satisfied with $m \leq j$. □

Now, define ${}^{\Pi}\mathbb{G}_\infty$ to be the subset of ${}^{\Pi}\mathbb{F}$ comprising all internal functions Φ satisfying (2) for some number $m \in \mathbb{N}_0$ and every compact $\mathbb{K}$ of $\mathbb{R}$ with $\mathbf{C}_{\mathbb{K},\Phi}$ a bounded positive constant. ${}^{\Pi}\mathbb{G}_\infty$ is a Π-difference algebra which contains ${}^{\Pi}\mathbb{D}_\infty$. Within ${}^{\Pi}\mathbb{G}_\infty$ the ordinary product of Π-predistributions make sense although the product of two Π-predistributions is not generally a Π-predistribution. By imposing appropriate restrictions on the factors, however, the product of two elements in ${}^{\Pi}\mathbb{D}_\infty$ may still be a Π-predistribution. In particular, we have

Theorem 2.3. *Let* $\Theta, \Phi \in {}^{\Pi}\mathbb{D}_\infty$ *be such that* $\mathbf{D}^m_+\Theta \in \mathbf{SC}_\Pi$ *and* $\Phi \in \mathbf{D}^m_+(\mathbf{SC}_\Pi)$ *for some given* $m \in \mathbb{N}_0$. *Then,* $\Phi = \mathbf{D}^m_+F$ *with* $F \in \mathbf{SC}_\Pi$, *and*

$$\mathrm{st}_{\mathcal{D}}\left(\Theta\Phi - \mathbf{D}^m_+\left(\sum_{j=0}^{m}\binom{m}{j}(-1)^j\mathbf{S}^{m-j}_+[(\mathbf{D}^{m-j}_+\Theta)F]\right)\right) = 0,$$

where $G \equiv \sum_{j=0}^{m} \binom{m}{j} (-1)^j \mathbf{S}_+^{m-j}[(\mathbf{D}_+^{m-j}\Theta)F]$ *is a function in* $\mathbf{SC}_\Pi$.

Proof. For $m = 1$ we have

$$\Theta \mathbf{D}_+ F = \mathbf{D}_+(\Theta F) - (\mathbf{D}_+\Theta)F - \varepsilon(\mathbf{D}_+\Theta)(\mathbf{D}_+F)$$

and therefore

$$\mathrm{st}_{\mathcal{D}}(\Theta\Phi - \mathbf{D}_+[\Theta F - \mathbf{S}_+(\mathbf{D}_+\Theta)F]) = \mathrm{st}_{\mathcal{D}}(-\varepsilon(\mathbf{D}_+\Theta)(\mathbf{D}_+F)).$$

For any $\varphi \in \mathcal{D}$, with support within the compact $\mathbb{K} \sqsubset \mathbb{R}$, we have that

$$\langle \mathrm{st}_{\mathcal{D}}(-\varepsilon(\mathbf{D}_+\Theta)(\mathbf{D}_+F)), \varphi \rangle = \mathrm{st}\left(\sum_{x \in {}^*\mathbb{K}_\Pi} \varepsilon\, \mathbf{D}_+\Theta(x)\, [F(x+\varepsilon) - F(x)] \right)$$

and the result follows from the fact that $F(x+\varepsilon) \approx F(x)$ for all $x \in {}^*\mathbb{K}_\Pi$. The proof now proceeds by induction on $m \in \mathbb{N}$. □

This result allow us to introduce the notion of Schwartz product in ${}^\Pi\mathcal{C}_\infty$ by setting

$$\Theta \cdot [\Phi] = [\Theta\Phi],$$

where Θ and Φ are as above.

2.2 *The* ${}^*\mathbb{C}_b$*-Module* ${}^\Pi\mathbb{D}$

For any subset A of $\mathbb{R}$ let $\kappa(A)$ be the family of all compact subsets of A. Denote by ${}^\Pi\mathbb{D}$ the subset of all functions $\Phi \in {}^\Pi\mathbb{F}$ such that for each $\mathbb{K} \in \kappa(\mathbb{R})$ there exist $\Phi_K \in {}^\Pi\mathbb{D}_\infty$ so that $\Phi = \Phi_K$ on ${}^*\mathbb{K}_\Pi$. Every function in ${}^\Pi\mathbb{D}$ determines a family

$$\{\Phi_\mathbb{K}\}_{\mathbb{K}\in\kappa(\mathbb{R})}$$

which is such that

$$\text{if } \mathbb{K}, \mathbb{L} \in \kappa(\mathbb{R}) \text{ and } \mathbb{K} \subset \mathbb{L} \text{ then } \Phi_\mathbb{K} = \Phi_\mathbb{L} \text{ on } {}^*\mathbb{K}_\Pi.$$

Such a family of ${}^\Pi\mathbb{D}_\infty$-functions is said to be *compatible*. Moreover the converse also holds, that is, if $\{\Phi_\mathbb{K}\}_{\mathbb{K}\in\kappa(\mathbb{R})}$ is a compatible family of internal functions in ${}^\Pi\mathbb{D}_\infty$ then we can define $\Phi \in {}^\Pi\mathbb{D}$ by setting

$$\Phi_{|_\mathbb{K}} = \Phi_\mathbb{K} \text{ on } {}^*\mathbb{K}_\Pi$$

for all $\mathbb{K} \in \kappa(\mathbb{R})$. Hence $\Phi \in {}^\Pi\mathbb{D}$.

If $\Phi \in {}^\Pi\mathbb{D}_\infty$ then the '*constant*' family $\{\Phi\}_{\mathbb{K}\in\kappa(\mathbb{R})}$ is certainly a compatible family and therefore defines an element in ${}^\Pi\mathbb{D}$; hence ${}^\Pi\mathbb{D}_\infty \subset {}^\Pi\mathbb{D}$. Every function in ${}^\Pi\mathbb{D}$ will be called a *global* Π-*predistribution*. Finite order Π-predistributions are global Π-predistributions, but the converse is not true, as the example that follows shows.

Example 2.4. Given the internal function

$$\Delta_0(x) = \begin{cases} \kappa & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

it is easy to see that for any $m \in {}^*\mathbb{N}_0$,

$$\mathbf{D}_+^m\Delta_0(x) = \begin{cases} (-1)^j\binom{m}{j}\kappa^{m+1} & \text{if } x = -(m-j)\varepsilon,\ j = 0, 1, \ldots, m \\ 0 & \text{otherwise.} \end{cases}$$

Moreover, for any $\varphi \in \mathcal{D}$, we get

$$^{\circ}\left(\sum_{x\in\Pi} \varepsilon\mathbf{D}_+^m\Delta_0(x)^*\varphi_\Pi(x)\right) = (-1)^m\varphi^{(m)}(0) = \langle\mathbf{D}^m\delta, \varphi\rangle$$

and hence Δ_0 is an hyperfinite representation of the (standard) delta distribution. $\mathbf{D}_+^m\Delta_0$ is, for every $m \in \mathbb{N}_0$, a function in $^\Pi\mathbb{D}_\infty$ and so is any finite linear combination (over $^*\mathbb{C}_b$) of these (finite order) Π-differences of Δ_0. However, the internal function

$$\Phi(x) = \sum_{n=-\kappa/2}^{\kappa/2-1} \mathbf{D}_+^{|n|}\Delta_0(x-n) \equiv \sum_{n=-\kappa/2}^{\kappa/2-1} \mathbf{D}_+^{|n|}\Delta_n(x)$$

is not in $^\Pi\mathbb{D}_\infty$ although, as it will be seen shortly, it belongs to $^\Pi\mathbb{D}$. To see this, note that for finite n the function $\mathbf{D}_+^{|n|}\Delta_0(x-n)$ is zero outside the Π-monad of n and for infinite n it is zero outside the Π-interval $[n-1/2, n+1/2]_\Pi$ which is completely contained in Π_∞. Thus for every compact $\mathbb{K} \in \kappa(\mathbb{R})$ the intersection of $^*\mathbb{K}_\Pi$ with the support of $\mathbf{D}_+^{|n|}\Delta_0(x-n)$ is empty, provided that $|n| \in {}^*\mathbb{N}_\infty$. Hence, for every $\mathbb{K} \in \kappa(\mathbb{R})$ there is only a finite number of finite-order Π-differences of finite-translates of Δ_0. Consequently, the restriction of Φ to $^*\mathbb{K}_\Pi$ is equal to a finite-order Π-difference of a function in $\mathbf{SC}_\Pi$.

The mapping $\mathrm{st}_\mathcal{D}$, defined on $^\Pi\mathbb{D}_\infty$, may now be extended to the whole of $^\Pi\mathbb{D}$ by setting

$$\mathrm{st}_\mathcal{D}(\Phi) = \{\mathrm{st}_{\mathcal{D}_\mathbb{K}}(\Phi_\mathbb{K})\}_{\mathbb{K}\sqsubset\mathbb{R}},$$

where $\mathrm{st}_{\mathcal{D}_\mathbb{K}}(\Phi_\mathbb{K})$ denotes the restriction of $\mathrm{st}_\mathcal{D}(\Phi_\mathbb{K})$ to $\mathcal{D}_\mathbb{K}$, for every $\mathbb{K} \in \kappa(\mathbb{R})$. That is to say, if $\varphi \in \mathcal{D}_\mathbb{K}$

$$\langle\mathrm{st}_\mathcal{D}(\Phi), \varphi\rangle = \mathrm{st}\left(\sum_{x\in^*\mathbb{K}_\Pi} \varepsilon\Phi_\mathbb{K}(x)^*\varphi_\Pi(x)\right).$$

$\mathrm{st}_\mathcal{D}$ is a linear map whose kernel, $\mathcal{K} \equiv \mathcal{K}(\mathrm{st}_\mathcal{D})$, comprises all internal functions in $^\Pi\mathbb{D}$ whose $\mathcal{D}$-standard part is the null distribution. Hence $^\Pi\mathbb{D}/\mathcal{K}$ is a linear space whose elements will be called *global* SΠ-*distributions.*

Note that for each $\mathbb{K} \in \kappa(\mathbb{R})$ there exist $m_\mathbb{K} \in \mathbb{N}_0$ and $F_\mathbb{K} \in \mathbf{SC}_\Pi$ such that $\Phi_\mathbb{K} = \mathbf{D}_+^{m_\mathbb{K}}F_\mathbb{K}$ on $^*\mathbb{K}_\Pi$. Thus, from Theorem 2 it follows that if $\Phi \in {}^\Pi\mathbb{D}$ then for every compact $\mathbb{K} \in \kappa(\mathbb{R})$ there exist a bounded positive constant $C_{\Phi,K}$ and an integer $m_K \in \mathbb{N}_0$, such that

$$\max_{x\in^*\mathbb{K}_\Pi} |\Phi(x)| \leq C_{\Phi,\mathbb{K}} \cdot \kappa^{m_\mathbb{K}}. \tag{3}$$

Define $^\Pi\mathbb{G}$ to be the set of all functions $\Phi \in {}^\Pi\mathbb{F}$ which satisfy the following property: for every compact $\mathbb{K} \in \kappa(\mathbb{R})$ there exist an integer $m_K \in \mathbb{N}_0$ and a finite number $C_{\Phi,\mathbb{K}}$ so that (3) holds. $^\Pi\mathbb{G}$ is a Π-difference algebra which contains $^\Pi\mathbb{D}$ as a linear submodule and $^\Pi\mathbb{G}_\infty$ as a subalgebra. Global Π-predistributions may therefore be multiplied within $^\Pi\mathbb{G}$. The product of two global Π-predistributions in general will not be a global

Π-predistribution. However, if $\Theta \in {}^{\Pi}\mathbb{D}$ is an internal function such that $\mathbf{D}_+^j\Theta \in \mathbf{SC}_\Pi$ for all (finite) $j \in \mathbb{N}_0$, and if $\Phi \in {}^{\Pi}\mathbb{D}$ then $\Theta\Phi$ is a global Π-predistribution in the sense that

$$\mathrm{st}_{\mathcal{D}_{\mathbb{K}}}\left(\Theta\Phi - \mathbf{D}_+^{m_K}\left(\sum_{j=0}^{m_K}\binom{m_K}{j}(-1)^j\mathbf{S}_+^{m_K-j}[(\mathbf{D}_+^{m_K-j}\Theta)F]\right)\right) = 0$$

for all compact $\mathbb{K} \sqsubset \mathbb{R}$. Hence, we define the product $\Theta[\Phi]$ to be the global Π-distribution $[\Theta\Phi]$.

3. The Π-Fourier transform

If F is a function in ${}^{\Pi}\mathbb{F}$ then, for each $y \in \Pi$, the sum

$$\hat{F}(y) = \sum_{x\in\Pi} \varepsilon^* \exp_\Pi(-2\pi ixy)F(x) \tag{4}$$

is a well-defined hypercomplex number. Thus, the right-hand side of (4) defines, for every $F \in {}^{\Pi}\mathbb{F}$, the internal function $\hat{F} : \Pi \longrightarrow {}^*\mathbb{C}$ which is also in ${}^{\Pi}\mathbb{F}$. Conversely, after some easy manipulations, we obtain

$$F(x) = \sum_{y\in\Pi} \varepsilon^* \exp_\Pi(2\pi ixy)\hat{F}(y) \tag{5}$$

which allows us to recover F from $\hat{F}$.

DEFINITION 3.1

Given $F \in {}^{\Pi}\mathbb{F}$, the function $\hat{F} \in {}^{\Pi}\mathbb{F}$, defined by (4), is called the Π-*Fourier transform* of F. Conversely F, as given by (5), is called the *inverse* Π-*Fourier transform* of $\hat{F}$.

Denoting the Π-Fourier transforms by $\mathcal{F}_\Pi$ and $\bar{\mathcal{F}}_\Pi$, respectively, then $\hat{F} = \mathcal{F}_\Pi[F]$ and $F = \bar{\mathcal{F}}_\Pi[\hat{F}]$. $\mathcal{F}_\Pi$ and $\bar{\mathcal{F}}_\Pi$ are linear transformations of ${}^{\Pi}\mathbb{F}$ onto ${}^{\Pi}\mathbb{F}$ and, moreover, $\mathcal{F}_\Pi \circ \bar{\mathcal{F}}_\Pi = \bar{\mathcal{F}}_\Pi \circ \mathcal{F}_\Pi = \mathrm{id}$.

Nonstandard hyperfinite versions for many of the properties of the (standard) Fourier transform and its inverse may be obtained. In particular, for any function $F \in {}^{\Pi}\mathbb{F}$, we obtain

$$\mathcal{F}_\Pi[\mathbf{D}_+F](y) = \sum_{x\in\Pi} \varepsilon^* \exp_\Pi(-2\pi ixy)\mathbf{D}_+F(x) = [-\lambda(y)]\hat{F}(y)$$

and, more generally, for any $j \in {}^*\mathbb{N}_0$,

$$\mathcal{F}_\Pi[\mathbf{D}_+^jF](y) = [-\lambda(y)]^j\hat{F}(y), \tag{6}$$

where $\lambda : \Pi \longrightarrow {}^*\mathbb{C}$ is the internal function defined by

$$\lambda(y) = \frac{1}{\varepsilon}[{}^*\exp_\Pi(2\pi i\varepsilon y) - 1]$$

and which is such that $\lambda(y) \approx 2\pi i(\mathrm{st}\ y)$ for every $y \in \Pi_b$. Also, for any $j \in {}^*\mathbb{N}_0$, we get

$$\mathbf{D}_+^j\hat{F}(y) = \mathcal{F}_\Pi[\bar{\lambda}^jF](y) \tag{7}$$

and, therefore, from (6) and (7), by inversion we obtain

$$\lambda^j(x)F(x) = \mathcal{F}_\Pi[\mathbf{D}^j_+\hat{F}](x), \tag{8}$$

$$\mathbf{D}^j_+F(x) = \bar{\mathcal{F}}_\Pi[(-\lambda)^j F](x). \tag{9}$$

Let $F, G \in {}^\Pi\mathbb{F}$ be any two internal function. Then, by simple manipulation, we may obtain the equation

$$\sum_{x\in\Pi} \varepsilon F(x)\hat{G}(x) = \sum_{y\in\Pi} \varepsilon \hat{F}(y)G(y) \tag{10}$$

which is the Π-*Parseval formula* in ${}^\Pi\mathbb{F}$.

3.1 *The* Π-*Fourier transform as an extension of the classical Fourier transform*

The (standard) classical Fourier transform is defined on $\mathbf{L}^1$, the space of all Lebesgue integrable functions on $\mathbb{R}$, by the integral

$$\mathcal{F}[f](\omega) = \int_{\mathbb{R}} f(t)\mathrm{e}^{-2\pi i\omega t}\mathrm{d}t, \quad \omega \in \mathbb{R}.$$

We denote by $\mathcal{F}_0$ the restriction of that transformation to $C_0 \cap \mathbf{L}^1 \subset \mathbf{L}^1$, the subspace of all continuous and integrable functions on $\mathbb{R}$ which tend monotonically to zero at infinity. Now we want to show that $\mathcal{F}_\Pi$ is an extension of $\mathcal{F}_0$ in the following sense:

Theorem 3.2. *For every* $f \in C_0 \cap \mathbf{L}^1$ *the equality*

$$\mathcal{F}_0[f](\mathrm{st}\, y) = \mathrm{st} \circ \mathcal{F}_\Pi[{}^\star f_\Pi](y)$$

holds for all $y \in \Pi_b$.

Proof. For any (fixed) $\omega \in \mathbb{R}$, let y be an arbitrarily given point in $\mathrm{st}_\Pi^{-1}(\omega)$. Defining for every $t \in \mathbb{R}$

$$f_y(t) = f(t)\exp[-2\pi i(\mathrm{st}\, y)t]$$

and extending this function to $\bar{\mathbb{R}}$ so that $f_y(\pm\infty) = 0$, consider the (external) function $f_y \circ \mathrm{st}_\infty(x), x \in \Pi$ (where $\mathrm{st}_\infty x = \mathrm{st}\, x$ if $x \in \Pi_b$ and $\mathrm{st}_\infty x = \pm\infty$ if $x \in \Pi^\pm_\infty$, respectively). Then we have that

$$\mathcal{F}_0[f](\omega) = \int_{\mathbb{R}} f_y(t)\mathrm{d}t = \int_\Pi f_y \circ \mathrm{st}_\infty(x)\mathrm{d}\Lambda_L(x),$$

where the last integral is the Loeb integral with respect to the Loeb counting measure on the hyperfinite grid. The proof will be complete provided it is shown that the equality

$$\int_\Pi f_y \circ \mathrm{st}_\infty(x)\mathrm{d}\Lambda_L(x) = \mathrm{st}\left(\sum_{x\in\Pi} \varepsilon^\star f_\Pi(x)^\star \exp_\Pi(-2\pi ixy)\right) \tag{11}$$

holds for all $y \in \Pi_b$. For this purpose it is necessary to prove that the internal function

$${}^\star f_\Pi(x)^\star \exp_\Pi(-2\pi ixy)$$

is an SΠ-integrable lifting for the external function $f_y \circ \mathrm{st}_\infty(x)$, $x \in \Pi$.

First, we have that

$$\mathrm{st}\{{}^\star f_\Pi(x)^\star \exp_\Pi(-2\pi ixy)\} = \mathrm{st}({}^\star f_\Pi(x)) \cdot \mathrm{st}({}^\star \exp_\Pi(-2\pi ixy))$$

and therefore:

- if $x \in \Pi_b$ then, from the continuity of the functions f and 'exp', it follows that
$$\mathrm{st}\{{}^\star f_\Pi(x)^\star \exp_\Pi(-2\pi ixy)\} = f(\mathrm{st}\,x)\exp[-2\pi i(\mathrm{st}\,x)(\mathrm{st}\,y)] = f_y \circ \mathrm{st}_\infty(x).$$
- if $x \in \Pi_\infty$ then since $f \in \mathcal{C}_0 \cap \mathbf{L}^1$ we have ${}^\star f_\Pi(x) \approx 0$; moreover the function ${}^\star \exp_\Pi(-2\pi ixy)$ is finitely bounded and therefore
$$\mathrm{st}\{{}^\star f_\Pi(x)^\star \exp_\Pi(-2\pi ixy)\} = f(\mathrm{st}\,x)\exp[-2\pi i(\mathrm{st}\,x)(\mathrm{st}\,y)] = 0 = f_y \circ \mathrm{st}_\infty(x).$$

Now it remains to show that the internal function ${}^\star f_\Pi(x)^\star \exp_\Pi(-2\pi ixy)$ is, for every (fixed) $y \in \Pi_b$, an SΠ-integrable function, that is, satisfies the following requirements:

(a) $\sum_{x \in \Pi_0^+} \varepsilon|{}^\star f_\Pi(x)|$ is finite,

(b) if $A \subset \Pi$ is internal and $\Lambda(A) \approx 0$ then $\sum_{x \in A} \varepsilon|{}^\star f_\Pi(x)| \approx 0$,

(c) if $A \subset \Pi$ is internal and ${}^\star f_\Pi(x) \approx 0, \forall_{x\in A}$ then $\sum_{x \in A} \varepsilon|{}^\star f_\Pi(x)| \approx 0$.

Since ${}^\star f_\Pi(x)$ is finitely bounded, taking into account that

$$\left|\sum_{x \in A} \varepsilon^\star f_\Pi(x)\right| \leq \sum_{x \in A} \varepsilon|{}^\star f_\Pi(x)| \leq \left\{\max_{x \in A} |{}^\star f_\Pi(x)|\right\} \cdot \Lambda(A)$$

shows that (b) follows immediately. We proceed now by proving the following lemma:

Lemma 3.3. *The hyperfinite* Π-*sum*

$$\sum_{|\gamma_1| \leq x \leq |\gamma_2|} \varepsilon|{}^\star f_\Pi(x)|$$

is infinitesimal for every two remote points $\gamma_1, \gamma_2 \in \Pi_\infty^+$ *(or, alternatively,* $\gamma_1, \gamma_1 \in \Pi_\infty^-$*) with* $|\gamma_1| \leq |\gamma_2| < \kappa/2$.

Proof of Lemma 3.3. Without any loss of generality we may take γ_1 and γ_2 to belong to $\Pi_\infty^+ \cap {}^\star\mathbb{N}_\infty$. Then we have

$$\sum_{\gamma_1 \leq x < \gamma_2} \varepsilon|{}^\star f_\Pi(x)| = \sum_{j=\gamma_1\kappa}^{\gamma_2\kappa - 1} \varepsilon|{}^\star f_\Pi(j\varepsilon)| = \sum_{n=\gamma_1}^{\gamma_2}\left\{\sum_{m=0}^{\kappa-1} \varepsilon|{}^\star f_\Pi(x)|\right\}$$

and therefore, taking into account that $|{}^\star f_\Pi(x)|$ is monotonically decreasing, we obtain

$$\sum_{\gamma_1 \leq x < \gamma_2} \varepsilon|{}^\star f_\Pi(x)| \leq \sum_{n=\gamma_1}^{\gamma_2} |{}^\star f_\Pi(x)|\left\{\sum_{m=0}^{\kappa-1} \varepsilon\right\} = \sum_{n=\gamma_1}^{\gamma_2} |{}^\star f_\Pi(n)|.$$

From the integral test it follows that the (standard) series

$$\sum_{n=1}^{\infty} |f(n)|$$

and the (standard) integral

$$\int_{1}^{+\infty} |f| \mathrm{d}\lambda$$

both converge or both diverge. Since the integral, by the hypothesis, is convergent, then the series also converges and therefore, from the nonstandard Cauchy convergence criterion for series, it follows that the hyperfinite sum

$$\sum_{n=\gamma_1}^{\gamma_2} |{}^*f_\Pi(n)|$$

is infinitesimal. ■

Now, for any arbitrarily fixed real number $e > 0$, define

$$\mathcal{N}_e = \left\{ n \in {}^*\mathbb{N} : \sum_{|j|=n\kappa}^{\frac{\kappa^2}{2}-1} \varepsilon |{}^*f_\Pi(j\varepsilon)| < e \right\}.$$

From Lemma 3.3, $\mathcal{N}_e$ contains arbitrarily small infinite numbers; since $\mathcal{N}_e$ is internal, then by underflow it contains a finite number, say $n_e \in \mathbb{N}$. That is,

$$\forall_n \left[n \in {}^*\mathbb{N} \wedge n_e \le n \le \kappa/2 \Rightarrow \sum_{|j|=n\kappa}^{\frac{\kappa^2}{2}-1} \varepsilon |{}^*f_\Pi(j\varepsilon)| < e \right].$$

Hence, since we have

$$\sum_{x\in\Pi} \varepsilon |{}^*f_\Pi(x)| = \sum_{|j|=0}^{n_e\kappa-1} \varepsilon |{}^*f_\Pi(j\varepsilon)| + \sum_{|j|=n_e\kappa}^{\frac{\kappa^2}{2}-1} \varepsilon |{}^*f_\Pi(j\varepsilon)| < \sum_{|j|=0}^{n_e\kappa-1} \varepsilon |{}^*f_\Pi(j\varepsilon)| + e$$

and, moreover,

$$\sum_{|j|=0}^{n_e\kappa-1} \varepsilon |{}^*f_\Pi(j\varepsilon)| \le n_e \left\{ \max_{-n_e \le x \le n_e} |{}^*f_\Pi(x)| \right\} < +\infty,$$

then (a) follows.

To prove (c) we reason as follows: (i) if $\Lambda(A)$ is finite, then the result follows from the fact that

$$\sum_{x\in A} \varepsilon |{}^*f_\Pi(x)| = \left\{ \max_{x\in A} |{}^*f_\Pi(x)| \right\} \cdot \Lambda(A) \approx 0;$$

(ii) if $\Lambda(A)$ is not finite then A certainly contains an infinite point in $\Pi^{\pm}_{\infty}$. Again from lemma 3.3 it follows that for any real $e > 0$ there exists (standard) $n_e \in \mathbb{N}$ such that

$$\sum_{|x|\in A\cap[n_e, \kappa/2-\varepsilon]} \varepsilon |{}^*f_\Pi(x)| < e,$$

while

$$\sum_{|x|\in A\cap[0,n_e-\varepsilon]} \varepsilon|{}^\star f_\Pi(x)| \approx 0.$$

Thus,

$$\sum_{x\in A} \varepsilon|{}^\star f_\Pi(x)| < e,$$

and, since $e > 0$ is arbitrary, the proof of (c) is complete.

Taking into account the definition of the (external) function $f \circ \mathrm{st}_\infty$, to prove the equality sign in (11) we need yet to show that the equality

$$\mathrm{st}\left(\sum_{x\in\Pi} \varepsilon^\star f_\Pi(x)\mathrm{e}^{-2\pi ixy}\right) = \mathrm{st}\left(\sum_{x\in\Pi} \varepsilon^\star f_\Pi(x)\mathrm{e}^{-2\pi ix(\mathrm{st}\,y)}\right)$$

holds. For this it is enough to show that the internal function

$$\hat{F}(y) = \sum_{x\in\Pi} \varepsilon^\star f_\Pi(x)\mathrm{e}^{-2\pi ixy}$$

is SΠ-continuous on Π_b. For $y, y' \in \Pi_b$ we have that

$$|\hat{F}(y) - \hat{F}(y')| \leq \sum_{x\in\Pi} \varepsilon|{}^\star f_\Pi(x)|\,|1 - \mathrm{e}^{-2\pi ix(y-y')}|.$$

From the fact that $f \in \mathcal{C}_0 \cap \mathbf{L}^1$ we have that, given a real number $r > 0$, the subset

$$\{x \in \Pi_b : |{}^\star f_\Pi(x)| < r/3\}$$

contains arbitrarily small infinite points; since the set is internal then by underflow there exists $x_r \in \Pi_b^+$ such that

$$\forall_x[x \in \Pi \wedge |x| > x_r \Rightarrow |{}^\star f_\Pi(x)| < r/3].$$

Then

$$\begin{aligned}|\hat{F}(y) - \hat{F}(y')| &\leq \left\{\sum_{|x|\leq x_r} + \sum_{|x|>x_r}\right\}\varepsilon|{}^\star f_\Pi(x)|\,|1 - \mathrm{e}^{-2\pi ix(y-y')}| \\ &\leq \sum_{|x|\leq x_r} \varepsilon|{}^\star f_\Pi(x)|\,|1 - \mathrm{e}^{-2\pi ix(y-y')}| + 2\sum_{|x|>x_r} \varepsilon|{}^\star f_\Pi(x)| \\ &< \frac{2r}{3} + \sum_{|x|\leq x_r} \varepsilon|{}^\star f_\Pi(x)|\,|1 - \mathrm{e}^{-2\pi ix(y-y')}|.\end{aligned}$$

Now, if $y \approx y'$ and x is finite then $2\pi ix(y - y') \approx 0$ and thus

$$\begin{aligned}&\sum_{|x|\leq x_r} \varepsilon|{}^\star f_\Pi(x)|\,|1 - \mathrm{e}^{-2\pi ix(y-y')}| \\ &\quad\leq \left\{\max_{|x|\leq x_r} |1 - \mathrm{e}^{-2\pi ix(y-y')}|\right\} \sum_{|x|\leq x_r} \varepsilon|{}^\star f_\Pi(x)| \approx 0 < \frac{r}{3}.\end{aligned}$$

Hence

$$|\hat{F}(y) - \hat{F}(y')| < r$$

and, since this is true for all real $r > 0$, it follows that

$$\forall_{y,y'}[y, y' \in \Pi_b \wedge y \approx y' \Rightarrow \hat{F}(y) \approx \hat{F}(y')]$$

that is, $\hat{F}$ is SΠ-continuous on Π_b and we have

$$\hat{f}(\operatorname{st} y) = \operatorname{st}(\hat{F}(y)), \quad y \in \Pi_b.$$

The proof is thereby complete. □

Given $f \in \mathcal{C}_0 \cap \mathbf{L}^1$ we may therefore obtain the Fourier transform of f by

$$\hat{f}(\operatorname{st} y) = \operatorname{st} \hat{F}(y), \quad y \in \Pi_b$$

where $\hat{F}$ is an SΠ-continuous function over Π. Hence $\hat{f}(\operatorname{st} y)$ is a continuous (and even uniformly continuous) function. Moreover, for every $y \in \Pi$,

$$|\hat{F}(y)| \leq \sum_{x\in\Pi} \varepsilon|{}^\star f_\Pi(x)|$$

which, since the right-hand side is finite, allow us to conclude that $\hat{F}(y)$, $y \in \Pi$ and $\hat{f}(\operatorname{st} y)$, $y \in \Pi_b$ are bounded functions.

The function $\hat{f}$, in general, does not belong to $\mathbf{L}^1$ and therefore, the inverse Fourier transform as defined by

$$\bar{\mathcal{F}}_\Pi[{}^\star\hat{f}_\Pi](x) = \sum_{y\in\Pi} \varepsilon{}^\star\hat{f}_\Pi(y){}^\star \exp_\Pi(2\pi ixy)$$

in general, does not allow us to recover the original function ${}^\star f_\Pi$ (and therefore f). For this purpose we have to take the inverse Π-Fourier transform of the function $\hat{F} = \mathcal{F}_\Pi[{}^\star f_\Pi]$

$$\begin{aligned}\bar{\mathcal{F}}_\Pi[{}^\star\hat{f}_\Pi](x) &\neq \sum_{y\in\Pi} \varepsilon\hat{F}(y){}^\star \exp_\Pi(2\pi ixy)\\ &= \hat{\mathcal{F}}_\Pi[\hat{F}](x) = {}^\star f_\Pi(x), \quad x \in \Pi.\end{aligned}$$

However, a nonstandard version of the Parseval's formula involving two functions $f, g \in \mathcal{C} \cap \mathbf{L}^1$, of the form

$$\sum_{y\in\Pi} \varepsilon{}^\star\hat{f}_\Pi(y){}^\star g_\Pi(y) \approx \sum_{x\in\Pi} \varepsilon{}^\star f_\Pi(x){}^\star\hat{g}_\Pi(x) \tag{12}$$

can be derived. Note that this is not the Π-Parseval's formula (10). To prove (12) it is enough to show that

$$\sum_{x\in\Pi} \varepsilon{}^\star f_\Pi(x)\hat{G}(x) \approx \sum_{x\in\Pi} \varepsilon{}^\star f_\Pi(x){}^\star\hat{g}_\Pi(x),$$

$$\sum_{y\in\Pi} \varepsilon\hat{F}(y){}^\star g_\Pi(y) \approx \sum_{y\in\Pi} \varepsilon{}^{\star\star}\hat{f}_\Pi(y){}^\star g_\Pi(y).$$

We will prove, for example, the second one since the other may be obtained similarly. Consider therefore

$$\sum_{x\in\Pi} \varepsilon\{\hat{F}(y) - {}^\star\hat{f}_\Pi(y)\}{}^\star g_\Pi(y).$$

For any $y \in \Pi_b$ we have that

$$^\star\hat{f}_\Pi(y) \approx \hat{f}(\mathrm{st}\, y) \approx \hat{F}(y)$$

and therefore the set

$$\{y \in \Pi : |y| > 0 \wedge |\hat{F}(y) - {}^\star\hat{f}_\Pi(y)| < 1/|y|\}$$

is internal and contains all finite $y \in \Pi_b$, $y \neq 0$. By overflow it contains $\eta \in \Pi_\infty^+$ such that

$$\forall_y[y \in \Pi \wedge |y| \leq \eta \Rightarrow \hat{F}(y) \approx {}^\star\hat{f}_\Pi(y)].$$

This fact, however, does not imply that the difference $\hat{F}(y) - {}^\star\hat{f}_\Pi(y)$ is kept at an infinitesimal level over the whole of the hyperfinite grid. Nevertheless, we have

$$\begin{aligned}\left|\sum_{y\in\Pi} \varepsilon\{\hat{F}(y) - {}^\star\hat{f}_\Pi(y)\}{}^\star g_\Pi(y)\right| &\leq \left|\left\{\sum_{|y|\leq\eta} + \sum_{|y|>\eta}\right\}\varepsilon\{\hat{F}(y) - {}^\star\hat{f}_\Pi(y)\}{}^\star g_\Pi(y)\right| \\ &= \left\{\max_{|y|\leq\eta} |\hat{F}(y) - {}^\star\hat{f}_\Pi(y)|\right\} \sum_{|y|\leq\eta} \varepsilon|{}^\star g_\Pi(y)| \\ &+ \left\{\max_{|y|>\eta} |\hat{F}(y) - {}^\star\hat{f}_\Pi(y)|\right\} \sum_{|y|>\eta} \varepsilon|{}^\star g_\Pi(y)|. \qquad (13)\end{aligned}$$

Now, because $g \in \mathcal{C}_0 \cap \mathbf{L}^1$, then

$$\sum_{|y|\leq\eta} \varepsilon|{}^\star g_\Pi(y)| \leq \int_{\mathbb{R}} |g(t)|\mathrm{d}t < +\infty$$

and

$$\sum_{|y|>\eta} \varepsilon|{}^\star g_\Pi(y)| \approx 0.$$

Moreover, $\max_{|y|\leq\eta}|\hat{F}(y) - {}^\star\hat{f}_\Pi(y)| \approx 0$ and $\max_{|y|>\eta}|\hat{F}(y) - {}^\star\hat{f}_\Pi(y)|$ is finite. Using all these facts in (13) we obtain finally

$$\sum_{y\in\Pi} \varepsilon\{\hat{F}(y) - {}^\star\hat{f}_\Pi(y)\}{}^\star g_\Pi(y) \approx 0$$

as asserted.

References

[1] Grenier J-P, Representation Discrete des Distributions Standard, *Osaka J. Math.* **32** (1995) 799–815

[2] Hoskins R F and Sousa-Pinto J, A nonstandard realization of the J. S. Silva axiomatic theory of distributions, *Port. Math.* **48**, nº 2 (1991), pp. 195–216

[3] Kinoshita Moto-o, Non-Standard Representations of Distributions I, *Osaka J. Math.* **25** (1988) 805–824

[4] Kinoshita Moto-o, Non-Standard Representations of Distributions II, *Osaka J. Math.* **27** (1990) 843–861

[5] Lindstrøm T, Nonstandard Analysis and its Applications (ed.) N Cutland, Students Text nº 10 (London Mathematical Society) (1988)

[6] Luxemburg W A, *Contributions to Nonstandard Analysis* (North Holland) (1972) pp. 15–39

Proc. Indian Acad. Sci. (Math. Sci.), Vol. 110, No. 4, November 2000, pp. 379–392.
© Printed in India

Sampling and Π-sampling expansions

J SOUSA PINTO and R F HOSKINS*

Departamento de Matemática, Universidade de Aveiro, Aveiro, Portugal
* Department of Mathematical Sciences, De Montfort University, Leicester, UK

MS received 28 February 2000

Abstract. Using the hyperfinite representation of functions and generalized functions this paper develops a rigorous version of the so-called 'delta method' approach to sampling theory. This yields a slightly more general version of the classical WKS sampling theorem for band-limited functions.

Keywords. Sampling expansions; WKS sampling theorem; non-standard analysis; hyperfinite sums.

1. Preliminaries

The classical sampling expansion for band-limited functions can be derived rigorously by several distinct arguments, but the use of so-called 'delta-methods' offers an approach which is intuitively most satisfying. A rigorous form of a delta-method derivation of the sampling expansion has been presented, using standard analysis, by Nashed and Walter [1]. In this paper we consider instead a non-standard approach to sampling theory. The hyperfinite representation of functions and generalized functions has been studied in an earlier paper [2], and the same notation and conventions will be used here. In particular, $\kappa \in {}^*\mathbb{N}_\infty$ denotes a given even infinite hypernatural number, $\varepsilon = \kappa^{-1} \approx 0$ and

$$\begin{aligned}\Pi \equiv \Pi_\kappa &= \left\{-\frac{\kappa}{2}, -\frac{\kappa}{2}+\varepsilon, \ldots, 0, \ldots, \frac{\kappa}{2}-\varepsilon\right\}\\ &= \left\{\left(-\frac{\kappa^2}{2}+j-1\right)\varepsilon : j = 1, 2, \ldots, \kappa^2\right\} \subset {}^*\mathbb{R}\end{aligned}$$

is the (unbounded) *hyperfinite line*. Given a standard point $r \in \mathbb{R}$, define the Π-*monad* of r by

$$\mathrm{mon}_\Pi(r) = \mathrm{st}_\Pi^{-1}(r) = \mathrm{mon}(r) \cap \Pi$$

where 'mon' denotes the usual monad of a standard number in ${}^*\mathbb{R}$. Then the set $\Pi_\mathrm{b} = \bigcup_{r\in\mathbb{R}} \mathrm{mon}_\Pi(r) = \mathrm{st}_\Pi^{-1}(\mathbb{R}) \subset \Pi$ is the *nearstandard hyperfinite line* and $\Pi_\infty = \Pi\backslash\Pi_\mathrm{b}$ is the set of *remote points* of the hyperfinite line. For every subset $A \subset \mathbb{R}$ define ${}^*A_\Pi = {}^*A \cap \Pi$ and $\mathrm{ns}_\Pi({}^*A) = {}^*A \cap \Pi_\mathrm{b} = \bigcup_{a\in A} \mathrm{mon}_\Pi(a)$.

By $\mathbb{F}_\Pi$ we denote the algebra of all internal functions $F : \Pi \longrightarrow {}^*\mathbb{C}$ which are periodically extended to the infinite grid $\varepsilon \cdot {}^*\mathbb{Z}$. The two difference operators $\mathbf{D}_+, \mathbf{D}_-$: $\mathbb{F}_\Pi \to \mathbb{F}_\Pi$ defined, for every function F and $x \in \Pi$, by

$$\mathbf{D}_+F(x) = \varepsilon^{-1}[F(x+\varepsilon) - F(x)] \quad \text{and} \quad \mathbf{D}_-F(x) = \varepsilon^{-1}[F(x) - F(x-\varepsilon)]$$

are called, respectively, the forward and the backward Π-difference operators (of first order). Iterating $\mathbf{D}_+$ (or $\mathbf{D}_-$) we obtain higher order Π-difference operators: for every (finite or infinite) $n \in {}^*\mathbb{N}_0$

$$\mathbf{D}_+^n F(x) = \mathbf{D}_+(\mathbf{D}_+^{n-1}F(x)), \quad x \in \Pi$$

and similarly for $\mathbf{D}_-^n$. It is easily seen that for any two functions $F, G \in \mathbb{F}_\Pi$ we have, (both for $\mathbf{D}_+$ and $\mathbf{D}_-$),

$$\mathbf{D}(F+G) = \mathbf{D}F + \mathbf{D}G \quad \text{and} \quad \mathbf{D}(F \cdot G) = (\mathbf{D}F)G + F(\mathbf{D}G) \pm \varepsilon(\mathbf{D}F)(\mathbf{D}G),$$

where we take $\pm\varepsilon$ according to the use of $\mathbf{D}_+$ or $\mathbf{D}_-$, respectively.

For every $\alpha, x \in \Pi$ define the Π-intervals (containing only points in Π) $J_\alpha^+(x)$ and $J_\alpha^-(x)$ as follows:

$$J_\alpha^+(x) = \begin{cases} (x,\alpha]_\Pi & \text{if } x < \alpha \\ [x,\alpha)_\Pi & \text{if } x > \alpha \end{cases}$$
$$J_\alpha^-(x) = \begin{cases} [\alpha,x)_\Pi & \text{if } x < \alpha \\ (\alpha,x]_\Pi & \text{if } x > \alpha \end{cases}$$

while for $x = \alpha$ we have $J_\alpha^+(x) = \emptyset = J_\alpha^-(x)$. For any $F \in \mathbb{F}_\Pi$ define the functions $\mathbf{S}_+F$ and $\mathbf{S}_-F$ to be the forward and backward Π-sums of F which are zero at the origin and which, for every $x \in \Pi\backslash\{0\}$ are defined by

$$\mathbf{S}_+F(x) = \sum_{t \in J_0^+(x)} \varepsilon F(t) \quad \text{and} \quad \mathbf{S}_-F(x) = \sum_{t \in J_0^-(x)} \varepsilon F(t).$$

The Π-*sum* operators $\mathbf{S}_+$ and $\mathbf{S}_-$ both transform $\mathbb{F}_\Pi$ into $\mathbb{F}_\Pi$. Moreover, for every $F \in \mathbb{F}_\Pi$ we have

$$\mathbf{D}_+\mathbf{S}_+F = F \quad \text{and} \quad \mathbf{D}_-\mathbf{S}_-F = F$$

that is, $\mathbf{S}_+$ and $\mathbf{S}_-$ are left inverses for $\mathbf{D}_+$ and $\mathbf{D}_-$, respectively.

Define the translation operator $\tau_\alpha : \mathbb{F}_\Pi \longrightarrow \mathbb{F}_\Pi$ (with $\alpha \in \Pi$) by setting $\tau_\alpha F(x) = F(x-\alpha)$ for every function F and $x \in \Pi$.

2. Π-periodic functions and Π-Fourier sums

2.1 Π-*Fourier sums*

For any internal function $F \in \mathbb{F}_\Pi$ define the Π-*periodic transform* of F with period 1 (or simply, the Π-periodic transform[1] of F) to be the internal function $\mathbf{T}_\Pi[F]$ in $\mathbb{F}_\Pi$ which is such that

$$\mathbf{T}_\Pi[F](x) = \sum_{n \in {}^*\mathbf{Z}_\Pi} F(x-n), \quad x \in \Pi$$

where ${}^*\mathbf{Z}_\Pi \equiv {}^*\mathbb{Z} \cap \Pi$. (As usual we suppose that the function $\mathbf{T}_\Pi[F]$ is periodically extended to the whole of the discrete line $\varepsilon^*\mathbb{Z}$.) In particular for the function Δ_0 defined by

$$\Delta_0(x) = \begin{cases} \kappa & \text{if } x = 0 \\ 0 & \text{if } x \neq 0 \end{cases}$$

[1] Unless explicitly stated, Π-periodic transforms will always be understood here to have period 1

we obtain

$$\mathbf{T}_\Pi[\Delta_0](x) = \sum_{n\in {}^*\mathbf{Z}_\Pi} \Delta_0(x-n)$$

which is the internal function in $\mathbb{F}_\Pi$ defined by

$$\mathbf{T}_\Pi[\Delta_0](x) = \begin{cases} \kappa & \text{if } x \in {}^*\mathbf{Z}_\Pi \\ 0 & \text{otherwise in } \Pi \end{cases}.$$

Considering the Π-convolution of an arbitrary function $F \in \mathbb{F}_\Pi$ with the function $\mathbf{T}_\Pi[\Delta_0]$, we get

$$\begin{aligned} F * \mathbf{T}_\Pi[\Delta_0](x) &= \sum_{y\in\Pi} \varepsilon F(x-y)\,\mathbf{T}_\Pi[\Delta_0](y) \\ &= \sum_{y\in\Pi} \varepsilon F(x-y)\left\{\sum_{n\in {}^*\mathbf{Z}_\Pi} \Delta_0(y-n)\right\} \\ &= \sum_{n\in {}^*\mathbf{Z}_\Pi}\left\{\sum_{y\in\Pi} \varepsilon F(x-y)\Delta_0(y-n)\right\} = \sum_{n\in {}^*\mathbf{Z}_\Pi} F(x-n) \end{aligned}$$

and, therefore, we may write

$$\mathbf{T}_\Pi[F] = F * \mathbf{T}_\Pi[\Delta_0]. \tag{1}$$

Taking the Π-Fourier transform of the internal function $\mathbf{T}_\Pi[\Delta_0]$, we obtain

$$\begin{aligned} \mathbf{F}_\Pi[\mathbf{T}_\Pi[\Delta_0]](y) &= \sum_{x\in\Pi} \varepsilon \mathbf{T}_\Pi[\Delta_0](x)\mathrm{e}^{2\pi ixy} = \sum_{x\in\Pi} \varepsilon \left\{\sum_{n\in {}^*\mathbf{Z}_\Pi} \Delta_0(x-n)\right\}\mathrm{e}^{2\pi ixy} \\ &= \sum_{n=-\kappa/2}^{\kappa/2-1}\left\{\sum_{x\in\Pi} \varepsilon\Delta_0(x-n)\mathrm{e}^{2\pi ixy}\right\} \\ &= \sum_{n=-\kappa/2}^{\kappa/2-1} \mathrm{e}^{2\pi iny} = \sum_{n=0}^{\kappa-1} \mathrm{e}^{2\pi i(n-\kappa/2)y} \\ &= \mathrm{e}^{-i\kappa\pi y}\sum_{n=0}^{\kappa-1} \mathrm{e}^{2\pi iny} = \frac{1-\mathrm{e}^{2i\pi\kappa y}}{1-\mathrm{e}^{2i\pi y}}\,{}^*\mathrm{exp}_\Pi(-i\pi\kappa y). \end{aligned}$$

Then, since $y = -\frac{\kappa}{2} + j\varepsilon$ where $j = 0, 1, \ldots, \kappa^2 - 1$, this sum vanishes for all y such that $j \neq r\kappa$ and has the value κ for all y such that $j = r\kappa$, where $r = 0, 1, \ldots, \kappa - 1$. Thus, we get

$$\mathbf{T}_\Pi[\Delta_0](y) = \mathbf{F}_\Pi[\mathbf{T}_\Pi[\Delta_0]](y) = \sum_{n\in {}^*\mathbf{Z}_\Pi} \mathrm{e}^{2\pi iny}, \quad y \in \Pi \tag{2}$$

which will be referred to as the Π-*Fourier sum*[2] for the internal function $\mathbf{T}_\Pi[\Delta_0]$.

[2] Similarly, we would obtain $\mathbf{T}_\Pi[\Delta_0](x) = \bar{\mathbf{F}}_\Pi[\mathbf{T}_\Pi[\Delta_0]](x) = \sum_{n\in {}^*\mathbf{Z}_\Pi} \mathrm{e}^{-2\pi inx}, \quad x \in \Pi.$

For an arbitrary internal function $F \in \mathbb{F}_\Pi$ and any $x \in \Pi$ we have

$$\begin{aligned}\mathbf{T}_\Pi[F](x) = F * \mathbf{T}_\Pi[\Delta_0](x) &= \sum_{y\in\Pi} \varepsilon F(x-y)\left\{\sum_{n\in{}^*\mathbf{Z}_\Pi} \mathrm{e}^{2\pi i n y}\right\} \\ &= \sum_{n\in{}^*\mathbf{Z}_\Pi}\left\{\sum_{y\in\Pi} \varepsilon F(x-y)\mathrm{e}^{2\pi i n y}\right\} \\ &= \sum_{n\in{}^*\mathbf{Z}_\Pi} \mathrm{e}^{2\pi i n x}\left\{\sum_{y\in\Pi} \varepsilon F(y)\mathrm{e}^{2\pi i n y}\right\} = \sum_{n\in{}^*\mathbf{Z}_\Pi} \hat{F}(n)\mathrm{e}^{2\pi i n x},\end{aligned}$$

where

$$\hat{F}(n) = \sum_{y\in\Pi} \varepsilon F(y)\mathrm{e}^{-2\pi i n y}, \quad n \in{}^*\mathbf{Z}_\Pi.$$

Thus

$$\mathbf{T}_\Pi[F](x) = \sum_{n\in{}^*\mathbf{Z}_\Pi} \hat{F}(n)\mathrm{e}^{2\pi i n x}, \quad x \in \Pi \tag{3}$$

is the Π-*Fourier sum* for the internal function $\mathbf{T}_\Pi[F]$. If, in particular, we take $F = \Delta_0$, then $\hat{\Delta}_0(n) = 1$ for all $n \in{}^*\mathbf{Z}_\Pi$ and we recover (2).

Writing (3) in the form

$$\sum_{n\in{}^*\mathbf{Z}_\Pi} F(x+n) = \sum_{n\in{}^*\mathbf{Z}_\Pi} \hat{F}(n)\mathrm{e}^{2\pi i n x}, \quad x \in \Pi$$

and setting $x = 0$ we obtain

$$\sum_{n\in{}^*\mathbf{Z}_\Pi} F(n) = \sum_{n\in{}^*\mathbf{Z}_\Pi} \hat{F}(n) \tag{4}$$

which is the Π-*Poisson formula* for the internal function $F(x), x \in \Pi$.

2.2 Π-*periodic functions*

An internal function $P \in \mathbb{F}_\Pi$ is said to be Π-*periodic* of period 1 (or, simply, Π-periodic)[3] if and only if

$$P(x+1) = P(x)$$

for all $x \in \Pi$. We denote by $\mathbb{P}_\Pi$ the subset of all Π-periodic internal functions in $\mathbb{F}_\Pi$.

Let F be any internal function in $\mathbb{F}_\Pi$. Since for any $x \in \Pi$ we have that

$$\begin{aligned}\mathbf{T}_\Pi[F](x+1) &= \sum_{n\in{}^*\mathbf{Z}_\Pi} F((x+1)-n) \\ &= \sum_{n\in{}^*\mathbf{Z}_\Pi} F(x-(n-1)) = \sum_{n\in{}^*\mathbf{Z}_\Pi} F(x-n) = \mathbf{T}_\Pi[F](x)\end{aligned}$$

then the Π-periodic transform of any internal function $F \in \mathbb{F}_\Pi$ belongs to $\mathbb{P}_\Pi$. Conversely, we have

[3] By Π-periodic functions we will always understand Π-periodic functions of period 1 defined on Π and periodically extended to the whole $\varepsilon^*\mathbb{Z}$.

Theorem 2.1. *Every internal function P in $\mathbb{P}_\Pi$ is the* Π*-periodic transform of an internal function Φ_P supported in* ${}^*[-1/2,+1/2)_\Pi$.[4]

Proof. Let $\mathbf{H}(x)$, $x \in \mathbb{R}$, denote the (usual) Heaviside unit step function defined by

$$\mathbf{H}(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}.$$

Then we have that

$${}^*\mathbf{H}_\Pi(x) = \sum_{t=-\frac{\kappa}{2}}^{x} \varepsilon\Delta_0(t)$$

and, moreover,

$${}^*\mathbf{H}_\Pi(x+1/2) - {}^*\mathbf{H}_\Pi(x-1/2), \quad x \in \Pi$$

is an internal function supported in the Π-interval ${}^*[-1/2, 1/2)_\Pi$. Hence,

$$\Phi_P(x) = [{}^*\mathbf{H}_\Pi(x+1/2) - {}^*\mathbf{H}_\Pi(x-1/2)]P(x), \quad x \in \Pi$$

defines an internal function in $\mathbb{F}_\Pi$ supported in ${}^*[-1/2,+1/2)_\Pi$, and

$$\begin{aligned}\mathbf{T}_\Pi[\Phi_P](x) &= \sum_{n \in {}^*\mathbf{Z}_\Pi} \Phi_P(x-n) \\ &= \sum_{n \in {}^*\mathbf{Z}_\Pi} [{}^*\mathbf{H}_\Pi(x+1/2-n) - {}^*\mathbf{H}_\Pi(x-1/2-n)]P(x-n) \\ &= P(x) \sum_{n \in {}^*\mathbf{Z}_\Pi} [{}^*\mathbf{H}_\Pi(x+1/2-n) - {}^*\mathbf{H}_\Pi(x-1/2-n)] = P(x)\end{aligned}$$

as stated. □

From this proposition it follows that

$$\begin{aligned}P(x) &= \mathbf{T}_\Pi[\Phi_P](x) = \Phi_P * \mathbf{T}_\Pi[\Delta_0](x) \\ &= \sum_{y \in \Pi} \varepsilon\Phi_P(y)\mathbf{T}_\Pi[\Delta_0](x-y) = \sum_{n \in {}^*\mathbf{Z}_\Pi} \mathrm{e}^{2\pi i n x}\left\{\sum_{y \in \Pi} \varepsilon\Phi_P(y)\mathrm{e}^{-2\pi i n y}\right\} \\ &= \sum_{n \in {}^*\mathbf{Z}_\Pi} \mathrm{e}^{2\pi i n x}\left\{\sum_{-1/2 \leq y < 1/2} \varepsilon\Phi_P(y)\mathrm{e}^{-2\pi i n y}\right\} = \sum_{n \in {}^*\mathbf{Z}_\Pi} c_{P,n}\mathrm{e}^{2\pi i n x},\end{aligned}$$

where, for every $n \in {}^*\mathbf{Z}_\Pi$,

$$c_{P,n} = \sum_{-1/2 \leq y < 1/2} \varepsilon\Phi_P(y)\mathrm{e}^{-2\pi i n y} = \hat{\Phi}_P(n)$$

is the nth Π-*Fourier coefficient* of the Π-periodic function $P \in \mathbb{P}_\Pi$. Hence

$$P(x) = \sum_{n \in {}^*\mathbf{Z}_\Pi} \hat{\Phi}_P(n)\mathrm{e}^{2\pi i n x}, \quad x \in \Pi$$

[4] That is to say, a function which is zero outside ${}^*[-1/2, 1/2)_\Pi$.

is the Π-*Fourier sum* of the internal function $P \in \mathbb{P}_\Pi$. Moreover, restricting suitably the variable x, we get

$$\Phi_P(x) = \sum_{n\in{}^*\mathbf{Z}_\Pi} \hat{\Phi}_P(n)\,e^{2\pi inx}, \quad -1/2 \le x < 1/2.$$

2.2.1 *The* Π-*Fourier sum as an extension of the classical Fourier series for continuous periodic functions*: Let f be a periodic function (with period 1), defined on the real line. Then there exists a function $g : [-1/2, 1/2) \to \mathbb{C}$ such that

$$f = \sum_{n=-\infty}^{+\infty} \tau_n \circ g.$$

The Π-nonstandard extension of g

$$\Phi(x) = {}^*g_\Pi(x)$$

is a finite SΠ-continuous function on ${}^*[-1/2, 1/2)_\Pi$ and, moreover, $\mathbf{T}_\Pi[\Phi]$ is an SΠ-continuous Π-periodic internal function. Then

$$\mathbf{T}_\Pi[\Phi](x) = \sum_{n\in{}^*\mathbf{Z}_\Pi} \hat{G}(n) e^{2\pi inx}, \tag{5}$$

where

$$\hat{G}(n) = \sum_{-1/2\le x<1/2} \varepsilon{}^*g_\Pi(x) e^{-2\pi inx}.$$

Suppose in addition that g is twice differentiable in $[-1/2, 1/2)$. Thus $\mathbf{D}_\Pi^2\Phi$ is always finite and, therefore, the sum

$$\Gamma(|\mathbf{D}_+^2\Phi|) \equiv \sum_{-1/2\le x<1/2} \varepsilon|\mathbf{D}_\Pi^2\Phi(x)|$$

is also finite. Hence we have that

$$\mathbf{F}_\Pi[\mathbf{D}_+^2\Phi](x) = \lambda(x)^2\hat{G}(x)$$

and so, for $n \neq 0$, we get

$$\hat{G}(n) = \frac{1}{|\lambda(n)|^2}|\mathbf{F}_\Pi[\mathbf{D}_\Pi^2\Phi](n)|.$$

Since $\lambda(n) = (2\pi in)\, e^{i\pi\varepsilon n}\, {}^*\mathrm{sinc}_\Pi(\varepsilon n)$ then, for every $n \in{}^*\mathbf{Z}_\Pi$, we have that $|\lambda(n)|^2 \ge 16|n|^2$. On the other hand, for any $n \in{}^*\mathbf{Z}_\Pi$, we have

$$|\bar{\mathbf{F}}_\Pi[\mathbf{D}_+^2\Phi](n)| = \left|\sum_{-1/2\le x<1/2} \varepsilon\mathbf{D}_+^2\Phi(x)e^{-2\pi inx}\right| \le \Gamma(|\mathbf{D}_+^2\Phi|),$$

where $\Gamma(|\mathbf{D}_+^2\Phi|)$ is a finite number. Hence, for large values of $|n|$, $n \in{}^*\mathbf{Z}_\Pi$, it follows that

$$|\hat{G}(n)| \leq \frac{\Gamma(|\mathbf{D}_+^2\Phi|)}{16} \cdot \frac{1}{|n|^2} \equiv C \cdot \frac{1}{|n|^2},$$

where C is a finite constant.

Let ν_1 and ν_2 be two infinite hypernatural numbers such that $\nu_1 \leq \nu_2 < \kappa/2$. Then, for any $x \in \Pi$, we have

$$\left|\sum_{|n|=\nu_1}^{\nu_2} \hat{G}(n)e^{2\pi inx}\right| \leq \sum_{|n|=\nu_1}^{\nu_2} |\hat{G}(n)| \leq 2C\sum_{n=\nu_1}^{\nu_2} \frac{1}{n^2}.$$

Since we have that

$$\sum_{n=\nu_1}^{\nu_2} \frac{1}{n^2} \approx 0$$

for all infinite $\nu_1, \nu_2 \in {}^*\mathbb{N}_\infty$, $(\nu_1 \leq \nu_2)$, it follows that

$$\left|\sum_{|n|=\nu_1}^{\nu_2} \hat{G}(n)e^{2\pi inx}\right| \approx 0$$

for all $x \in \Pi_b$ and all $\nu_1, \nu_2 \in {}^*\mathbb{N}_\infty$, $(\nu_1 \leq \nu_2)$.

Then we have

Theorem 2.2. *If t denotes any (standard) point in $[-1/2, 1/2) \subset \mathbb{R}$, then the Π-Fourier sum (5) reads as follows*

$$f(t) = \mathrm{st} \circ \mathbf{T}_\Pi[\Phi](x) = \mathrm{st}\left(\sum_{n=-\frac{\kappa}{2}}^{\frac{\kappa}{2}-1} \hat{G}(n)\, e^{2\pi inx}\right) = \sum_{n=-\infty}^{+\infty} \hat{g}(n)e^{2\pi int}$$

for all $x \in \mathrm{mon}_\Pi(t)$.

Proof. For each $x \in \mathrm{mon}_\Pi(t)$ and for every real $r > 0$, the set

$$\left\{\nu \in {}^*\mathbf{Z}_\Pi \cap {}^*\mathbb{N} : \left|\sum_{|n|=\nu}^{\frac{\kappa}{2}-1} \hat{G}(n)e^{2\pi inx}\right| < r\right\}$$

is internal and contains all infinite positive numbers in ${}^*\mathbf{Z}_\Pi$. Then, by underflow it contains a finite number $n_r \in \mathbb{N}$. Hence

$$\left|\sum_{n=-\frac{\kappa}{2}}^{\frac{\kappa}{2}-1} \hat{G}(n)e^{2\pi inx} - \sum_{n=-n_r}^{n_r} \hat{G}(n)e^{2\pi inx}\right| < r$$

and thus, taking standard parts, we obtain

$$\left|\mathrm{st}\left(\sum_{n=-\frac{\kappa}{2}}^{\frac{\kappa}{2}-1} \hat{G}(n)e^{2\pi inx}\right) - \sum_{n=-n_r}^{n_r} \mathrm{st}\,\hat{G}(n) \cdot \mathrm{st}\; e^{2\pi inx}\right| < r.$$

Taking into account that for $x \in \mathrm{mon}_\Pi(t)$ and $|n| \in \mathbb{N}_0$ we have $e^{2\pi inx} \approx e^{2\pi int}$ and $\mathrm{st}\,\hat{G}(n) = \hat{g}(n)$ then, since the real number $r > 0$ is arbitrary, it follows that

$$f(t) = \mathrm{st} \circ \mathbf{T}_\Pi[\Phi](x)$$
$$= \mathrm{st}\left(\sum_{n=-\frac{\kappa}{2}}^{\frac{\kappa}{2}-1} \hat{G}(n)e^{2\pi inx}\right) = \sum_{n=-\infty}^{+\infty} \hat{g}(n)e^{2\pi int}$$

for any $x \in \mathrm{mon}_\Pi(t)$. □

3. Π-Bandlimited functions

3.1 *Basic definitions and results*

Denote by $\mathbf{I}$ the unitary Π-interval $\mathbf{I} = \{-\frac{1}{2} \cdots \frac{1}{2} - \varepsilon\}$ and let the function $\hat{F}$ in $\mathbb{F}_\Pi$ be supported on $\mathbf{I}$ (that is, $\hat{F}(y) = 0$ for all $y \notin \mathbf{I}$). Then, the internal function $F = \bar{\mathbf{F}}_\Pi[\hat{F}] \in \mathbb{F}_\Pi$ is said to be Π-*bandlimited* to $\mathbf{I}$. Denote by $\mathbb{B}_\Pi \equiv \mathbb{B}_\Pi(\mathbf{I})$ the subspace of $\mathbb{F}_\Pi$ comprising all functions which are Π-bandlimited[5].

For any function $F \in \mathbb{B}_\Pi$ we have

$$F(x) = \sum_{y\in\Pi} \varepsilon\hat{F}(y)e^{2\pi ixy} = \sum_{-1/2\leq y<1/2} \varepsilon\hat{F}(y)e^{2\pi ixy}$$

and therefore the inequality

$$|F(x)| \leq \Gamma(|\hat{F}|)$$

holds for all $x \in \Pi$. The function F may extend to the hyperfinite plane $\Pi + i\Pi$ by setting

$$F(\xi + i\eta) = \sum_{-1/2\leq y<1/2} \varepsilon\,\hat{F}(y)e^{2\pi i(\xi+i\eta)y}$$
$$= \sum_{-1/2\leq y<1/2} \varepsilon\{e^{-2\pi\eta y}\hat{F}(y)\}e^{2\pi i\xi y}.$$

Hence we get

$$|F(\xi + i\eta)| \leq \Gamma(|\hat{F}|) \cdot \exp(\pi|\eta|)$$

for all $\xi + i\eta \in \Pi + i\Pi$.

Moreover, for any $j \in {}^\star\mathbb{N}_0$, we obtain

$$\mathbf{D}^j_+F(x) = \sum_{-1/2\leq y<1/2} \varepsilon\lambda^j(y)\hat{F}(y)e^{2\pi ixy}$$

and therefore

$$|\mathbf{D}^j_+F(x)| \leq C_j \cdot \Gamma(|\hat{F}|),$$

where

$$C_j = \max_{-1/2\leq y<1/2} |\lambda^j(y)| \leq 2\pi \max_{-1/2\leq y<1/2} |t|^j = 2^{1-j}\pi.$$

For finite j the constant C_j is finite; for infinite j the constant C_j is infinitesimal.

[5] We will consider here only functions Π-bandlimited to $\mathbf{I}$ and these will be referred to simply as Π-*bandlimited functions*. The generalization to other (finite or hyperfinite) intervals is immediate.

Hence, if $\Gamma(|\hat{F}|)$ is finite the function $F \equiv \bar{\mathbf{F}}_{\Pi}[\hat{F}]$ is finite and so also are its Π-derivatives over the whole hyperfinite line; moreover, F extends to $\Pi + i\Pi$ as a finite function of exponential type $\leq \pi$. In general F is not $S\Pi$-continuous on Π_{b} and therefore $\mathrm{st} \circ F$ may not exist; however, the equation

$$\begin{aligned}\langle F, {}^{\star}\hat{\varphi}_{\Pi}\rangle &= \sum_{x \in \Pi} \varepsilon F(x)^{\star}\hat{\varphi}_{\Pi}(x) \\ &\approx \sum_{y \in \mathbf{I}} \varepsilon \hat{F}(y)^{\star}\varphi_{\Pi}(y) = \langle \hat{F}, {}^{\star}\varphi_{\Pi}\rangle\end{aligned}$$

holds for any function φ such that $\varphi \circ \mathrm{st} = \mathrm{st} \circ {}^{\star}\varphi_{\Pi}$ and $\hat{\varphi} \circ \mathrm{st} = \mathrm{st} \circ {}^{\star}\hat{\varphi}_{\Pi}$. This is certainly true for any function φ which is continuous and supported on the interval $[-1/2, 1/2]$; therefore F defines a continuous linear functional over the linear space which is the Fourier transform of the space $\mathcal{C}[-1/2, 1/2]$ (with the topology generated by the uniform norm).

3.2 *The Π-sampling expansion*

Let F be a Π-bandlimited function in $\mathbb{B}_{\Pi}$. Then we have

$$F(x) = \sum_{y \in \Pi} \varepsilon \hat{F}(y) \mathrm{e}^{2\pi i x y} = \sum_{-1/2 \leq y < 1/2} \varepsilon \hat{F}(y) \mathrm{e}^{2\pi i x y} \tag{6}$$

for every $x \in \Pi$. Now, for $y \in \mathbf{I}$, we have

$$\hat{F}(y) = \mathbf{T}_{\Pi}[\hat{F}](y) = \sum_{n=-\frac{\kappa}{2}}^{\frac{\kappa}{2}-1} c_n \mathrm{e}^{-2\pi i n y},$$

where, for every $n \in {}^{\star}\mathbf{Z}_{\Pi}$, the coefficient c_n is given by

$$c_n = \sum_{-1/2 \leq y < 1/2} \varepsilon \hat{F}(y) \mathrm{e}^{2\pi i n y} = F(n).$$

Replacing $\hat{F}(y)$ for its Π-Fourier sum in (6), we obtain

$$\begin{aligned}F(x) &= \sum_{-1/2 \leq y < 1/2} \varepsilon \left\{ \sum_{n=-\frac{\kappa}{2}}^{\frac{\kappa}{2}-1} F(n) \mathrm{e}^{-2\pi i n y} \right\} \mathrm{e}^{2\pi i x y} \\ &= \sum_{n=-\frac{\kappa}{2}}^{\frac{\kappa}{2}-1} F(n) \left\{ \sum_{-1/2 \leq y < 1/2} \varepsilon \mathrm{e}^{2\pi i (x-n) y} \right\} = \sum_{n=-\frac{\kappa}{2}}^{\frac{\kappa}{2}-1} F(n) \Gamma(x-n),\end{aligned}$$

where $\Gamma : \Pi \to {}^{\star}\mathbb{C}$ is defined by

$$\begin{aligned}\Gamma(x) &= \sum_{-1/2 \leq y < 1/2} \varepsilon \mathrm{e}^{2\pi i x y} = \sum_{m=0}^{\kappa-1} \varepsilon\, \mathrm{e}^{2\pi i x (-\frac{1}{2} + m\varepsilon)} \\ &= \varepsilon\, \mathrm{e}^{-i\pi x} \sum_{m=0}^{\kappa-1} \mathrm{e}^{2\pi i \varepsilon x m} = \varepsilon \frac{1 - \mathrm{e}^{2\pi i x}}{1 - \mathrm{e}^{2\pi i \varepsilon x}} \mathrm{e}^{-i\pi x} = \varepsilon \frac{{}^{\star}\mathrm{sin}_{\Pi}(\pi x)}{{}^{\star}\mathrm{sin}_{\Pi}(\pi \varepsilon x)} \mathrm{e}^{-i\pi\varepsilon x}.\end{aligned}$$

Since

$$\lambda(x) = 2\pi ix\,\mathrm{e}^{i\pi\varepsilon x}\,{}^*\mathrm{sinc}_\Pi(\varepsilon x)$$

then, finally, we get

$$\Gamma(x) = \frac{2\pi ix}{\lambda(x)}\,{}^*\mathrm{sinc}_\Pi(x), \quad x \in \Pi. \tag{7}$$

We thus obtain

$$F(x) = \sum_{n=-\kappa/2}^{\kappa/2-1} F(n)\frac{2\pi i(x-n)}{\lambda(x-n)}\,{}^*\mathrm{sinc}_\Pi(x-n), \quad x \in \Pi \tag{8}$$

which will be called the Π-*sampling expansion* for the internal function $F \in \mathbb{B}_\Pi$. The internal function F is expressed in terms of its values at the hyperinteger points $n \in {}^*\mathbf{Z}_\Pi$ and the Π-*interpolating function* Γ.

We consider now some properties of the Π-interpolating function Γ. First, extending Γ by continuity, we have

$$\Gamma(0) = \frac{2\pi ix}{\lambda(x)}\,{}^*\mathrm{sinc}_\Pi(x)\Big|_{x=0} = 1. \tag{9}$$

Since the zeros of the function $\sin(\pi\varepsilon x)$ are of the form $x = j\kappa$, $j \in {}^*\mathbf{Z}$, then the function $\lambda(x)$ has no zeros inside the hyperfinite line Π other than $x = 0$. Hence $\Gamma(x)$ is a well-defined function on Π. Moreover we have that $\Gamma(x) = 0$ for $x = -\frac{\kappa}{2}, \ldots, -2, -1, 1, 2, \ldots, \frac{\kappa}{2} - 1$.

For any $\alpha \in \Pi$ such that $|\alpha| \le \frac{\pi}{2}$ we have[6] that $\frac{2}{\pi}|\alpha| \le |\sin\alpha| \le |\alpha|$; thus, since $|x| \le \kappa/2$, we obtain

$$\frac{2}{\pi}|\pi\varepsilon x| \le |\sin(\pi\varepsilon x)| \le |\pi\varepsilon x|$$

and therefore

$$4|x| \le |\lambda(x)| \le 2\pi|x|.$$

Hence $|(2\pi ix)/\lambda(x)| \le \pi/2$ and so

$$|\Gamma(x)| \le \pi/2, \quad \text{for all } x \in \Pi$$

that is, Γ is a finitely bounded internal function on Π. Moreover, since

$$\mathbf{D}^j_+\Gamma(x) = \sum_{-1/2 \le y < 1/2} \varepsilon\lambda^j(y)\mathrm{e}^{2\pi ixy}$$

and $\lambda^j(\cdot)\mathrm{e}^{2\pi ix\cdot}$ is SΠ-integrable on $\mathbf{I}$ for every $j \in \mathbb{N}_0$, then the function Γ is easily seen to be such that

$$\mathbf{D}^j_+\Gamma \text{ is S}\Pi\text{-continuous on } \Pi_\mathrm{b}$$

[6] See Abramowitz and Stagen, *Handbook of Mathematical Functions* (Dover) (1972).

for all $j \in \mathbb{N}_0$. Therefore, we may define the infinitely differentiable (standard) function $\gamma : \mathbb{R} \longrightarrow \mathbb{R}$, by setting for every $t \in \mathbb{R}$,

$$\gamma(t) = \operatorname{st}\Gamma(x),$$

for any $x \in \operatorname{mon}_\Pi(t)$. For $x + i\eta \in {}^*\mathbb{C}$ we have that

$$\Gamma(x+i\eta) = \sum_{-1/2 \leq y < 1/2} \mathrm{e}^{-2\pi\eta y}\mathrm{e}^{2\pi i x y}$$

and so

$$|\Gamma(x+i\eta)| \leq C\exp(\pi|\eta|).$$

Hence γ also extends into the complex plane as an entire function of exponential type $\leq \pi$. Since

$$\Gamma(x) = \frac{\mathrm{e}^{-i\pi\varepsilon x}}{{}^*\mathrm{sinc}_\Pi(\varepsilon x)}\, {}^*\mathrm{sinc}_\Pi(x)$$

and $\mathrm{e}^{i\pi\varepsilon x}/{}^*\mathrm{sinc}_\Pi(\varepsilon x) \approx 1$ for every $x \in \Pi_\mathrm{b}$, we have

$$\Gamma(x) \approx {}^*\mathrm{sinc}_\Pi(x)$$

for all $x \in \Pi_\mathrm{b}$; on the other hand,

$$\Gamma(x) \approx 0$$

for all infinite $x \in \Pi$. In fact,

- if $\varepsilon x = x/\kappa$ is infinitesimal, then

$$\Gamma(x) \approx {}^*\mathrm{sinc}_\Pi(x) \approx 0,$$

- if $\varepsilon x = x/\kappa$ is not infinitesimal then, taking into account the estimate for the function $|\lambda(x)|$, we obtain

$$|\Gamma(x)| \leq \frac{\pi}{2}\left|\frac{\sin(\pi x)}{\pi x}\right| \approx 0.$$

3.2.1 *The* Π*-sampling expansion as an extension of the sampling expansion for a* (*standard*) *bandlimited function with continuous and differentiable spectrum*: Let $\hat{f}$ be a continuous function with compact support within the (standard) interval $[-1/2, +1/2] \subset \mathbb{R}$. Then, the internal function

$$\hat{F}(y) = {}^*\hat{f}_\Pi(y), \quad -1/2 \leq y < 1/2$$

is finite, SΠ-continuous and such that $\operatorname{st}\hat{F} = \hat{f} \circ \operatorname{st}$; moreover, $\Gamma(|\hat{F}|)$ is certainly a finite number.

The inverse Π-Fourier transform of $\hat{F}$ is, itself, an internal function in $\mathbf{SC}_\Pi$. In fact, for $x, \xi \in \Pi_\mathrm{b}$, we have

$$\begin{aligned} F(x) - F(\xi) &= \sum_{-1/2 \leq y < 1/2} \varepsilon\hat{F}(y)\mathrm{e}^{2\pi i \xi y}\{\mathrm{e}^{2\pi i (x-\xi) y} - 1\} \\ &= 2\pi i(x-\xi) \sum_{-1/2 \leq y < 1/2} \varepsilon y \hat{F}(y)\mathrm{e}^{2\pi i \xi y}[1 + 2\pi i (x-\xi) y \mathbf{O}(1)] \end{aligned}$$

and therefore

$$\begin{aligned}|F(x) - F(\xi)| &\leq 2\pi|x-\xi| \cdot \sum_{y\in\mathbf{I}} \varepsilon|y\hat{F}(y)|\,|1 + 2\pi i(x-\xi)y\mathbf{O}(1)| \\ &= 2\pi|x-\xi|\left\{\max_{y\in\mathbf{I}}|1+2\pi i(x-\xi)y\mathbf{O}(1)|\right\}\sum_{y\in\mathbf{I}}\varepsilon|y\hat{F}(y)| \\ &= \pi|x-\xi|\left\{\max_{y\in\mathbf{I}}|1+2\pi i(x-\xi)y\mathbf{O}(1)|\right\}\Gamma(|\hat{F}|).\end{aligned}$$

Hence, if $x \approx \xi$ we have that $F(x) \approx F(\xi)$ on Π_b, as asserted. Thus the projection $f = \mathrm{st}\, F$ is a well-defined continuous function on $\mathbb{R}$ and is the inverse Fourier transform of the given function $\hat{f}$. Moreover, since

$$\mathbf{D}_+^j F(x) = \sum_{-1/2\leq y<1/2} \varepsilon\lambda^j(y)\hat{F}(y)\mathrm{e}^{2\pi ixy}$$

then it is easily seen that $\mathbf{D}_+^j F$ belongs to $\mathbf{SC}_\Pi$ for all $j \in \mathbb{N}_0$ (but not necessarily so for infinite $j \in {}^*\mathbb{N}_0$). Thus since

$$\mathrm{st}(\mathbf{D}_+^j F) = f^{(j)}$$

we may assert that $f \equiv \mathrm{st}\ F$ is a C^∞-function. Further, we have

$$F(x+i\eta) = \sum_{y\in\mathbf{I}} \varepsilon\{\mathrm{e}^{-2\pi\eta y}\hat{F}(y)\}\mathrm{e}^{2\pi ixy}$$

and so F extends finitely to $\Pi_b + i\Pi_b$ such that

$$|F(x+i\eta)| \leq C\exp(\pi|\eta|).$$

The projection $f = \mathrm{st}\ F$, therefore, extends over the whole finite complex plane as an entire function of exponential type $\leq \pi$.

Additionally suppose now that $\hat{f}$ is such that[7]

$$\Gamma(|\mathbf{D}_+\hat{F}|) = \sum_{y\in\mathbf{I}} \varepsilon|\mathbf{D}_+\hat{F}(y)|$$

is a finite number. For any infinite $n \in {}^*\mathbf{Z}_\Pi \cap {}^*\mathbb{Z}_\infty$ we have that

$$\bar{\lambda}(n)F(n) = \sum_{-1/2\leq y<1/2} \varepsilon\mathbf{D}_+\hat{F}(y)\mathrm{e}^{2\pi iny}$$

and therefore

$$|\bar{\lambda}(n)\,F(n)| \leq \Gamma(|\mathbf{D}_+\hat{F}|)$$

or, taking into account that $|\bar{\lambda}(n)| \geq 4|n|$, we obtain

$$|F(n)| \leq C\cdot\frac{1}{|n|},$$

[7] This condition certainly holds if the (standard) function $\hat{f}$ is differentiable everywhere on $[-\frac{1}{2},\frac{1}{2}] \subset \mathbb{R}$.

where $C = \Gamma(|\mathbf{D}_+ \hat{F}|)/4$. Now, for every $x \in \Pi_b$ and $n \in {}^*Z_\Pi \cap {}^*\mathbb{Z}_\infty$, we have that

$$\begin{aligned} |\Gamma(x-n)| &\leq \frac{\pi}{2}\left|\frac{\sin(\pi(x-n))}{\pi(x-n)}\right| \leq \frac{1}{2}\frac{1}{|x-n|} \\ &\leq \frac{1}{2}\frac{1}{||x|-|n||} \leq \frac{1}{2|n|}\frac{1}{1-|x/n|} \leq \frac{1}{|n|} \end{aligned}$$

since $x/n \approx 0$ and thus $1 - |x/n| \geq 1/2$. Hence, for $\nu_1, \nu_2 \in {}^*\mathbb{N}_\infty$ such that $\nu_1 \leq \nu_2 < \kappa/2$, we have

$$\left|\sum_{|n|=\nu_1}^{\nu_2} F(n)\Gamma(x-n)\right| \leq 2C \cdot \sum_{n=\nu_1}^{\nu_2} \frac{1}{n^2}.$$

On the other hand,

$$\sum_{n=\nu_1}^{\nu_2} \frac{1}{n^2} \approx 0$$

for all infinite $\nu_1, \nu_2 \in {}^*\mathbb{N}_\infty$, $(\nu_1 \leq \nu_2)$ and therefore

$$\sum_{|n|=\nu_1}^{\nu_2} F(n)\,\Gamma(x-n) \approx 0.$$

Then we have

Theorem 3.1. *Let f be any standard function whose Fourier transform $\hat{f}$ is a continuous and differentiable function on $[-\frac{1}{2}, \frac{1}{2}]$. If t denotes any (standard) point in $\mathbb{R}$, the Π-sampling expansion* (8) *reads as follows:*

$$f(t) = \operatorname{st} F(x) = \operatorname{st}\left(\sum_{n=-\frac{\kappa}{2}}^{\frac{\kappa}{2}-1} F(n)\Gamma(x-n)\right) = \sum_{n=-\infty}^{+\infty} f(n)\operatorname{sinc}(t-n),$$

where x is any point in $\operatorname{mon}_\Pi(t)$. Moreover, the convergence of the (standard) series on the right-hand side is almost uniform on $\mathbb{R}$ (that is to say is uniform on compacts).

Proof. If $\hat{f}$ is differentiable everywhere on $[-\frac{1}{2}, \frac{1}{2}]$ then, although $\operatorname{st} \circ \mathbf{D}_+\hat{F}$ may not be equal to $\hat{f}' \circ \operatorname{st}$, we certainly have that $\mathbf{D}_+\hat{F}$ is finite everywhere on $\mathbf{I}$ and therefore $\Gamma(|\mathbf{D}_+\hat{F}|)$ is a finite number. Hence, from above, for each $x \in \operatorname{mon}_\Pi(t)$ and every real $r > 0$, the set

$$\left\{\nu \in {}^*\mathbf{Z}_\Pi \cap {}^*\mathbb{N} : \left|\sum_{|n|=\nu}^{\frac{\kappa}{2}-1} F(n)\Gamma(x-n)\right| < r/2\right\}$$

is internal and contains all infinite positive numbers in ${}^*\mathbf{Z}_\Pi$. By underflow it contains a finite number, say $n_r \in \mathbb{N}$, for which we have

$$\left|\sum_{n=-\frac{\kappa}{2}}^{\frac{\kappa}{2}-1} F(n)\Gamma(x-n) - \sum_{n=-n_r}^{n_r} F(n)\Gamma(x-n)\right| < r/2.$$

Thus, taking standard parts, gives

$$\left|\operatorname{st}\left(\sum_{n=-\frac{\kappa}{2}}^{\frac{\kappa}{2}-1} F(n)\Gamma(x-n)\right) - \sum_{n=-n_r}^{n_r} \operatorname{st} F(n) \operatorname{st} \Gamma(x-n)\right| < r.$$

Taking into account that for $x \in \operatorname{mon}_\Pi(t)$ and $|n| \in \mathbb{N}_0$ we have $\Gamma(x-n) \approx \Gamma(t-n)$ and therefore

$$\operatorname{st}\Gamma(x-n) = \operatorname{st}\Gamma(t-n) = \operatorname{sinc}(t-n).$$

Since the real number $r > 0$ is arbitrary, it follows that

$$\begin{aligned} f(t) &= \operatorname{st} F(x) \\ &= \operatorname{st}\left(\sum_{n=-\frac{\kappa}{2}}^{\frac{\kappa}{2}-1} F(n)\Gamma(x-n)\right) = \sum_{n=-\infty}^{+\infty} f(n)\operatorname{sinc}(t-n) \end{aligned}$$

for any $x \in \operatorname{mon}_\Pi(t)$, where the rightmost hand side is to be interpreted in the standard sense. Moreover, since for all $x, z \in \Pi_b$ such that $x \approx z$ and for all $\nu \in {}^*\mathbf{Z}_\Pi \cap {}^*\mathbb{N}_\infty$, we have that

$$\sum_{|n|=\nu}^{\frac{\kappa}{2}-1} F(n)\Gamma(x-n) \approx \sum_{|n|=\nu}^{\frac{\kappa}{2}-1} F(n)\Gamma(z-n)$$

and so the series converges almost uniformly on $\mathbb{R}$. □

From the whole proof above it follows that this result also holds under the following slightly more general form:

Theorem 3.2. *Let f be any standard function whose Fourier transform $\hat{f}$ is continuous on $[-\frac{1}{2}, \frac{1}{2}]$ and such that $\Gamma(|\mathbf{D}_+\hat{F}|)$ is finite. If t denotes any (standard) point in $\mathbb{R}$, the Π-sampling expansion (8) reads as follows*

$$f(t) = \operatorname{st} F(x) = \operatorname{st}\left(\sum_{n=-\frac{\kappa}{2}}^{\frac{\kappa}{2}-1} F(n)\Gamma(x-n)\right) = \sum_{n=-\infty}^{+\infty} f(n)\operatorname{sinc}(t-n),$$

where x is any point in $\operatorname{mon}_\Pi(t)$. Moreover, the convergence of the (standard) series on the right-hand side is almost uniform on $\mathbb{R}$.

References

[1] Nashed M, Zuhair and Walter C G, General Sampling Theorems for Functions in Reproducing Kernel Hilbert Spaces, *Mathematics of Control, Signals and Systems* **4** (1991) 363–390

[2] Sousa-Pinto J and Hoskins R F, Hyperfinite Representation of Distributions, *Proc. Indian Acad. Sci.* **110** (2000) 363

Proc. Indian Acad. Sci. (Math. Sci.), Vol. 110, No. 4, November 2000, pp. 393–414.
© Printed in India

Formula for a solution of $u_t + H(u, Du) = g$

ADIMURTHI and G D VEERAPPA GOWDA

TIFR Centre, P.B. 1234, Indian Institute of Science Campus, Bangalore 560 012, India
E-mail: aditi@math.tifrbng.res.in; gowda@math.tifrbng.res.in

MS received 10 February 2000; revised 21 March 2000

Abstract. We study the continuous as well as the discontinuous solutions of Hamilton–Jacobi equation $u_t + H(u, Du) = g$ in $\mathbb{R}^n \times \mathbb{R}_+$ with $u(x, 0) = u_0(x)$. The Hamiltonian $H(s, p)$ is assumed to be convex and positively homogeneous of degree one in p for each s in $\mathbb{R}$. If H is non increasing in s, in general, this problem need not admit a continuous viscosity solution. Even in this case we obtain a formula for discontinuous viscosity solutions.

Keywords. Hamilton–Jacobi equation; dynamic programming principle; viscosity sub and super solutions.

1. Introduction

Let $H : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$ be a continuous function. Let $u_0 \in W^{1,\infty}(\mathbb{R}^n)$, $g \in W^{1,\infty}(\mathbb{R}^n \times \mathbb{R}_+)$ and consider the Hamilton–Jacobi equation

$$\begin{aligned} &u_t + H(u, Du) = g \quad &&\text{in } \mathbb{R}^n \times \mathbb{R}_+, \\ &u(x, 0) = u_0(x) &&\text{for } x \in \mathbb{R}^n, \end{aligned} \tag{1.1}$$

where $Du = \left(\frac{\partial u}{\partial x_1}, \dots, \frac{\partial u}{\partial x_n}\right)$. This problem has been studied extensively using the method of viscosity solutions developed by Crandall, Evans and Lions [7, 10]. An excellent reference for this is the lecture notes of Evans [7]. It has been shown ([10], Ch. 9, Remark 9.1) that apart from the usual hypothesis on H, if there exist a $\gamma \in \mathbb{R}$ such that for all $(s, p) \in \mathbb{R} \times \mathbb{R}^n$,

$$\frac{\partial H}{\partial s}(s, p) \geq \gamma, \tag{1.2}$$

then (1.1) admits a viscosity solution $u \in W^{1,\infty}(\mathbb{R}^n \times \mathbb{R}_+)$. The question is to obtain a formula for the solution of (1.1).

Under the conditions (1.2), $p \mapsto H(s, p)$ being convex and positively homogeneous of degree one and $g \equiv 0$, Barron, Jenssen and Liu [5] have obtained an explicit formula for the viscosity solution. In general for $g \not\equiv 0$, one cannot expect an explicit formula in the sense of Hopf and Lax, but one can hope to get an infinite dimensional representation in terms of a control problem. This has been carried out by Barron and Ishii [3] (representation formula) and Barron and Liu [4] (existence of a minimizer).

If H does not satisfy (1.2), in general (1.1) need not admit a continuous viscosity solution [1, 10]. In this paper we study this problem under the hypothesis, $s \mapsto H(s, p)$ is non-increasing and $p \mapsto H(s, p)$ is convex and positively homogeneous of degree one. As far as our knowledge goes this problem has not been tackled in the literature. But in [1], the authors considered this problem with $g = 0$ and obtained explicit formula for solutions. Here we extend this result for $g \not\equiv 0$ (see Theorem 2.2). The main ingredients in the proof of this are to prove semicontinuity property for the constraints (Corollary 4.1) and the dynamic programming principle (Lemmas 4.7 and 4.8). The same idea allows us to study problem (1.1) when H satisfies (1.2) (see Theorem 2.1) and of course this result can be obtained also from the results of [3] and [4] with proper modifications.

2. Main results

Let $x \in \mathbb{R}^n, t > 0, 0 \leq s < t, M > 0$ and define

$$C(x,t,s) = \{\xi \in W^{1,\infty}([s,t],\mathbb{R}^n);\ \xi(t) = x\}, \tag{2.1}$$

$$C_M(x,t,s) = \{\xi \in C(x,t,s);\ |\dot{\xi}|_\infty \leq M\}, \tag{2.2}$$

$$C(x,t) = C(x,t,0), C_M(x,t) = C_M(x,t,0),$$

where $\dot{\xi}(\theta) = \mathrm{d}\xi/\mathrm{d}\theta(\theta)$. Let $h : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ and $u : \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}$ be functions and $g \in W^{1,\infty}(\mathbb{R}^n \times \mathbb{R}_+)$. Let $0 \leq s \leq \theta \leq t$, $\xi \in C(x,t,s)$, define

$$|u|_t = \sup\{|u(y,\theta)|;\ y \in \mathbb{R}^n, \theta \in [0,t]\},$$

$$\int_s^\theta g(\xi) = \int_s^\theta g(\xi(\lambda),\lambda)\,\mathrm{d}\lambda, \tag{2.3}$$

$$\rho_+(\xi,t,s,h,g) = \operatorname*{ess\,sup}_{\theta\in[s,t]} \left\{h(\dot{\xi}(\theta)) - \int_s^\theta g(\xi)\right\}, \tag{2.4}$$

$$\rho_-(\xi,t,s,h,g) = \operatorname*{ess\,inf}_{\theta\in[s,t]} \left\{h(\dot{\xi}(\theta)) - \int_s^\theta g(\xi)\right\}, \tag{2.5}$$

$$\rho_\pm(\xi,t,h,g) = \rho_\pm(\xi,t,0,h,g). \tag{2.6}$$

Then we have the following results.

Theorem 2.1. *Let $u_0 \in W^{1,\infty}(\mathbb{R}^n)$, $g \in W^{1,\infty}(\mathbb{R}^n \times \mathbb{R}_+)$. Assume that H satisfies,*

(H_1) *$s \mapsto H(s,p)$ is non decreasing for all $p \in \mathbb{R}^n$.*

(H_2) *$p \mapsto H(s,p)$ is convex, positively homogeneous of degree one for each $s \in \mathbb{R}$.*

Let h denote the quasi convex dual of H defined by

$$h(q) = \inf\{\gamma : H(\gamma,p) \geq \langle p,q\rangle\ \forall\,|p| = 1\}. \tag{2.7}$$

For $(x,t) \in \mathbb{R}^n \times \mathbb{R}_+$, define

$$u(x,t) = \inf_{\xi\in C(x,t)} \left\{u_0(\xi(0)) \vee \rho_+(\xi,t,h,g) + \int_0^t g(\xi)\right\}. \tag{2.8}$$

Then for each $T > 0$, $u \in W^{1,\infty}(\mathbb{R}^n \times [0,T])$ and is a viscosity solution of (1.1). Furthermore infimum is achieved in (2.8).

Theorem 2.2. *Let $u_0 \in W^{1,\infty}(\mathbb{R}^n), g \in W^{1,\infty}(\mathbb{R}^n \times \mathbb{R}_+)$. Assume that H satisfies*

(H_3) *$s \mapsto H(s,p)$ is non-increasing for all $p \in \mathbb{R}^n$,*

(H_4) *$p \mapsto H(s,p)$ is convex, positively homogeneous of degree one, for each $s \in \mathbb{R}$.*

Let h denote the quasi concave dual of H defined by

$$h(q) = \sup\{\gamma : H(\gamma,p) \geq \langle p,q\rangle \ \forall\ |p| = 1\}. \tag{2.9}$$

For $(x,t) \in \mathbb{R}^n \times \mathbb{R}_+$, define

$$\underline{u}(x,t) = \inf_{\xi\in C(x,t)} \left\{u_0(\xi(0)) + \int_0^t g(\xi);\ u_0(\xi(0)) \leq \rho_-(\xi,t,h,g)\right\}, \tag{2.10}$$

$$\bar{u}(x,t) = \inf_{\xi\in C(x,t)} \left\{u_0(\xi(0)) + \int_0^t g(\xi);\ u_0(\xi(0)) < \rho_-(\xi,t,h,g)\right\}. \tag{2.11}$$

Then $\underline{u}$ is a lower semicontinuous viscosity solution of (1.1) and $\bar{u}$ is an upper semicontinuous viscosity solution of (1.1). Also infimum is achieved in (2.10). Furthermore if $g(x,t) = g_1(x,t) + g_2(t)$, $t \to tg_1$ is non-increasing in t, $g_2(t) \leq 0$ and $H(u,p) > 0$ for all $p \neq 0, u \in \mathbb{R}$, then $\underline{u}^* = \bar{u}$ and $\bar{u}_* = \underline{u}$. In this case the two solutions coincide.

Remark 2.3. For $g \equiv 0$, using Jenssen's inequality as in Lemma 3.3 of [5], Theorem 2.1 reduces to Theorem 3.1 of [5]. Also Theorem 2.2 reduces to Theorem 2.1 of [1].

3. Preliminaries

In this section we recall the definitions and known results from [9, 6, 5] and [4] without proofs.

DEFINITION 3.1

Let $\Omega \subset \mathbb{R}^n$ be a domain and V be a locally bounded function. For $x \in \overline{\Omega}$ define

$$V^*(x) = \limsup_{r\to 0}\{V(z) : |z-x| \leq r\},$$
$$V_*(x) = \liminf_{r\to 0}\{V(z) : |z-x| \leq r\}.$$

Then V^* is an upper semicontinuous and V_* is a lower semicontinuous functions and $V_* \leq V \leq V^*$.

As in [6] and [9], we have the following:

DEFINITION 3.2

Let U be a locally bounded function in $\mathbb{R}^n \times \mathbb{R}_+$.

1. U is said to be a subsolution of (1.1) if for any $(x_0,t_0) \in \mathbb{R}^n \times \mathbb{R}_+, \varphi \in C^1(\mathbb{R}^n \times \mathbb{R}_+)$ such that (x_0,t_0) is a local maximum for $U^* - \varphi$ with $U^*(x_0,t_0) = \varphi(x_0,t_0)$, then at (x_0,t_0), $\varphi_t + H(\varphi, D\varphi) \leq g$ and $U^*(x,0) \leq u_0(x)$.
2. U is said to be a super solution of (1.1) if for any $(x_0,t_0) \in \mathbb{R}^n \times \mathbb{R}_+, \varphi \in C^1(\mathbb{R}^n \times \mathbb{R}_+)$ such that (x_0,t_0) is a local minimum for $U_* - \varphi$ with $U_*(x_0,t_0) = \varphi(x_0,t_0)$, then at $(x_0,t_0), \varphi_t + H(\varphi, D\varphi) \geq g$ and $U_*(x,0) \geq u_0(x)$.
3. U is said to be a viscosity solution of (1.1) if U is both a sub and a super solution.

Now recall some properties of quasi convex (concave) dual of H.

Let H satisfy (H_1) and (H_2). Then

(A$_1$) h is a lower semicontinuous quasi convex function i.e, for any $q_1, q_2 \in \mathbb{R}^n, t \in [0,1]$, $h(tq_1 + (1-t)q_2) \leq \max\{h(q_1), h(q_2)\}$,

(A_2) $\inf h = -\infty, \lim_{|q|\to\infty} h(q) = \infty$,
(A_3) $H(s,p) = \sup\{\langle p,q\rangle; h(q) \le s\}$.

Proofs of (A_1) to (A_3) follow from lemmas (2.1) and (2.2) of [5].

Let H satisfy (H_3) and (H_4) of Theorem 2.2 and let h be the quasi concave dual of H. Then

(A_4) h is an upper semicontinuous quasi concave function i.e, for $t \in [0,1]$, $q_1, q_2 \in \mathbb{R}^n$, $h(tq_1 + (1-t)q_2) \ge \min\{h(q_1), h(q_2)\}$,
(A_5) $\sup h = +\infty, \lim_{|q|\to\infty} h(q) = -\infty$,
(A_6) $H(s,p) = \sup\{\langle p,q\rangle : s \le h(q)\}$.

(A_4) to (A_6) follow from (A_1) to (A_3) applied to the Hamiltonian $\tilde{H}(s,p) = H(-s,p)$.

4. Proof of theorems

Before going to the proof of the theorems, we need the following lemma for proving the existence of a minimizer and the semicontinuity of $\underline{u}$ and $\overline{u}$.

Let $0 \le s < t$ and $1 \le p \le \infty$. Let $b, b_\kappa : L^p([s,t]) \times [s,t] \to \mathbb{R}$ be continuous functions. Assume that b, b_κ satisfies the following hypothesis: For $\xi, \xi_\kappa \in L^p([s,t]), \theta$, $\theta_\kappa \in [s,t]$ such that $\xi_\kappa \to \xi$ strongly in L^p and $\theta_\kappa \to \theta$ as $\kappa \to \infty$, then

$$\lim_{\kappa\to\infty} b_\kappa(\xi_\kappa, \theta_\kappa) = b(\xi, \theta).$$

Lemma 4.1. *Let $h : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ be a function and b, b_κ satisfying the above hypothesis and $\eta, \eta_\kappa \in L^p([s,t], \mathbb{R}^n)$ such that $\eta_\kappa \rightharpoonup \eta$ weakly. Then*

(a) *Assume that h is a lower semicontinuous quasi convex function. Then*

$$\varliminf_{\kappa\to\infty} \operatorname*{ess\,sup}_{\theta\in[s,t]} \{h(\eta_\kappa(\theta)) + b_\kappa(\xi_\kappa, \theta)\} \ge \operatorname*{ess\,sup}_{\theta\in[s,t]} \{h(\eta(\theta)) + b(\xi, \theta)\} \tag{4.1}$$

(b) *Assume that h is an upper semicontinuous quasi concave function. Then*

$$\varlimsup_{\kappa\to\infty} \operatorname*{ess\,inf}_{\theta\in[s,t]} \{h(\eta_\kappa(\theta)) + b_\kappa(\xi_\kappa, \theta)\} \le \operatorname*{ess\,inf}_{\theta\in[s,t]} \{h(\eta(\theta)) + b(\xi, \theta)\} \tag{4.2}$$

Proof. Observe that (a) follows from (b) by changing h to $-h, b_\kappa$ to $-b_\kappa$. Hence it is enough to prove (b). Let

$$C = \varlimsup_{\kappa\to\infty} \operatorname*{ess\,inf}_{\theta\in[s,t]} \{h(\eta_\kappa(\theta)) + b_\kappa(\xi_\kappa, \theta)\}.$$

If $C = -\infty$, then there is nothing to prove. Let $m > 0$ and choose $\kappa(m)$ such that $\kappa(m) \to \infty$ as $m \to \infty$ and for any $\kappa \ge \kappa(m)$

$$C - \frac{1}{m} \le \operatorname*{ess\,inf}_{\theta\in[s,t]} \{h(\eta_\kappa(\theta)) + b_\kappa(\xi_\kappa, \theta)\}.$$

Since $\eta_\kappa \rightharpoonup \eta$ as $\kappa \to \infty$, hence there exist $0 \le \alpha_{\kappa l} \le 1, \Sigma_{\kappa\ge\kappa(m)} \alpha_{\kappa l} = 1, \alpha_{\kappa l \neq 0}$ for all but a finitely many κ such that $f_l = \Sigma_{\kappa\ge\kappa(m)} \alpha_{\kappa l} \eta_\kappa \to \eta$ strongly in L^p as $l \to \infty$, ([11], theorem 3.13). Hence extracting a subsequence still denoted by f_l and a null set $N \subset [s,t]$

such that for all $\theta \notin N$, $f_l(\theta) \to \eta(\theta)$. Let $\theta \notin N$ and choose a $\kappa_l \geq \kappa(m)$ such that

$$\min_{\alpha_{\kappa l} \neq 0} \{h(\eta_\kappa(\theta))\} = h(\eta_{\kappa_l}(\theta)).$$

Then by quasi concavity we have

$$\begin{aligned} h(f_l(\theta)) + b_{\kappa_l}(\xi_{\kappa_l}, \theta) &\geq \min_{\alpha_{\kappa_l} \neq 0} \{h(\eta_\kappa(\theta)) + b_{\kappa_l}(\xi_{\kappa_l}, \theta)\} \\ &= h(\eta_{\kappa_l}(\theta)) + b_{\kappa_l}(\xi_{\kappa_l}, \theta) \geq C - \frac{1}{m}. \end{aligned}$$

Since h is upper semicontinuous, now letting $l \to \infty$ and $m \to \infty$, we obtain for all $\theta \notin N$

$$C \leq h(\eta(\theta)) + b(\xi, \theta)$$

and this proves (4.2).

As a consequence of this lemma we have the following result. Let $0 \leq s_\kappa < t_\kappa, 0 \leq s < t$ such that $(s_\kappa, t_\kappa) \to (s, t)$ as $\kappa \to \infty$. Let $\{\xi_\kappa\} \in W^{1,\infty}([s_\kappa, t_\kappa], \mathbb{R}^n)$ be a bounded sequence and $g \in C^0(\mathbb{R}^n \times \mathbb{R}_+) \cap L^\infty$. Let $\alpha_\kappa : [s_\kappa, t_\kappa] \to [s, t]$ be defined by $\alpha_\kappa(\theta) = \frac{t-s}{t_\kappa - s_\kappa}\theta + \frac{t_\kappa s - t s_\kappa}{t_\kappa - s_\kappa}$, $\tilde{\xi}_\kappa(\theta) = \xi_\kappa(\alpha_\kappa^{-1}(\theta))$ and $b_\kappa(\tilde{\xi}_\kappa, \theta) = \frac{s_\kappa - t_\kappa}{t-s} \int_s^\theta g(\tilde{\xi}_\kappa(\lambda), \alpha_\kappa^{-1}(\lambda))\mathrm{d}\lambda$.

COROLLARY 4.1

Assume that as $\kappa \to \infty$, $\tilde{\xi}_\kappa \to \xi$ in C^0-topology for some $\xi \in W^{1,\infty}([s,t], \mathbb{R}^n)$. Let $h : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ be a function. Then

(a) *Assume that h is a lower semicontinuous and quasi convex function, then*

$$\varliminf_{\kappa \to \infty} \operatorname*{ess\,sup}_{\theta \in [s_\kappa, t_\kappa]} \left\{ h(\dot{\xi}_\kappa(\theta)) - \int_{s_\kappa}^\theta g(\xi_\kappa) \right\} \geq \operatorname*{ess\,sup}_{\theta \in [s,t]} \left\{ h(\dot{\xi}(\theta)) - \int_s^\theta g(\xi) \right\}. \tag{4.3}$$

(b) *Assume that h is an upper semicontinuous and quasi concave function, then*

$$\varlimsup_{\kappa \to \infty} \operatorname*{ess\,inf}_{\theta \in [s_\kappa, t_\kappa]} \left\{ h(\dot{\xi}_\kappa(\theta)) - \int_{s_\kappa}^\theta g(\xi_\kappa) \right\} \leq \operatorname*{ess\,inf}_{\theta \in [s,t]} \left\{ h(\dot{\xi}(\theta)) - \int_s^\theta g(\xi) \right\}. \tag{4.4}$$

Proof. (a) follows from (b) by changing h to $-h$ and g to $-g$. Hence it is enough to prove (b). By change of variables we have

$$\begin{aligned} &\operatorname*{ess\,inf}_{\theta \in [s_\kappa, t_\kappa]} \left\{ h(\dot{\xi}_\kappa(\theta)) - \int_{s_\kappa}^\theta g(\xi_\kappa) \right\} \\ &\quad = \operatorname*{ess\,inf}_{\theta \in [s,t]} \left\{ h\left(\left(\frac{t-s}{t_\kappa - s_\kappa} \right) \dot{\tilde{\xi}}_\kappa(\theta) \right) + b_\kappa(\tilde{\xi}_\kappa, \theta) \right\}. \end{aligned}$$

Since $\{\dot{\tilde{\xi}}_\kappa\}$ is a bounded sequence in $L^2([s,t], \mathbb{R}^n)$ and $\tilde{\xi}_\kappa \to \xi$ strongly in L^2, hence if $\eta_\kappa(\theta) = \frac{t-s}{t_\kappa - s_\kappa} \dot{\tilde{\xi}}_\kappa(\theta)$, then $\eta_\kappa \rightharpoonup \dot{\xi}$ weakly in L^2. Now (4.4) follows from (4.2). This proves the corollory.

Theorem 2.1.

In order to prove Theorem 2.1, we will first establish the dynamic programming principle. In order to do this we need some information on the bounds of the cost function.

Lemma 4.2. *Let W be a function on $\mathbb{R}^n \times \mathbb{R}_+$. Assume that for every $T > 0$, $|W|_T = \sup\{|W(x,t)|; (x,t) \in \mathbb{R}^n \times [0,T]\} < \infty$. For $0 \leq s < t \leq T$ and $x \in \mathbb{R}^n$, define*

$$V(x,t) = \inf_{\xi \in C(x,t,s)} \left\{ W(\xi(s),s) \vee \rho_+(\xi,t,s,h,g) + \int_s^t g(\xi) \right\}. \tag{4.5}$$

Then there exist a constant $M(T) > 0$ such that

$$|V(x,t)| \leq |W|_T + T|g|_\infty, \tag{4.6}$$

$$V(x,t) = \inf_{\xi \in C_{M(T)}(x,t,s)} \left\{ W(\xi(s),s) \vee \rho_+(\xi,t,s,h,g) + \int_s^t g(\xi) \right\}. \tag{4.7}$$

Furthermore if W is a continuous function, then there exist a $\xi \in C_{M(T)}(x,t,s)$ such that

$$V(x,t) = W(\xi(s),s) \vee \rho_+(\xi,t,s,h,g) + \int_s^t g(\xi). \tag{4.8}$$

Proof. Let $M_1(T) = |W|_T + T|g|_\infty$. Since $\inf h = -\infty$, there exist a $q \in \mathbb{R}^n$ such that $h(q) + T|g|_\infty < -|W|_T$. Let $\xi(\theta) = x + q(\theta - t) \in C(x,t,s)$ and hence $\rho_+(\xi,t,s,h,g) \leq h(q) + T|g|_\infty < -|W|_T$. Therefore $V(x,t) \leq |W|_T + T|g|_\infty \leq M_1(T)$. Also for any $\xi \in C(x,t,s)$,

$$W(\xi(s),s) \vee \rho_+(\xi,t,s,h,g) + \int_s^t g(\xi) \geq W(\xi(s),s) + \int_s^t g(\xi) \geq -|W|_T - T|g|_\infty.$$

Hence $|V(x,t)| \leq M_1(T)$. This proves (4.6).

Since $\lim_{|p|\to\infty} h(p) = \infty$, we can choose a $M(T) > 0$ such that whenever $|p| \geq M(T)$ then $h(p) \geq 3(|W|_T + T|g|_\infty)$. Let $\xi \in C(x,t,s)$ such that $|\dot{\xi}|_\infty \geq M(T)$. Then we have

$$\begin{aligned} \operatorname*{ess\,sup}_{\theta \in [s,t]} \left\{ h(\dot{\xi}(\theta)) - \int_s^\theta g(\xi) \right\} &\geq \operatorname*{ess\,sup}_{\theta \in [s,t]} \{h(\dot{\xi}(\theta))\} - T|g|_\infty \\ &\geq 3(|W|_T + T|g|_\infty) - T|g|_\infty \\ &\geq 2|W|_T. \end{aligned}$$

Hence from (4.6) we have

$$\begin{aligned} W(\xi(s),s) \vee \rho_+(\xi,t,s,h,g) + \int_s^t g(\xi) &= \rho_+(\xi,t,s,h,g) + \int_s^t g(\xi) \\ &\geq 3(|W|_T + T|g|_\infty) - 2T|g|_\infty \\ &> |V(x,t)|. \end{aligned}$$

This proves (4.7).

Let W be a continuous function. Since V is bounded, from (4.7) we can choose a sequence $\xi_\kappa \in C_{M(T)}(x,t,s)$ such that

$$V(x,t) = \lim_{\kappa\to\infty} \left\{ W(\xi_\kappa(s),s) \vee \rho_+(\xi_\kappa,t,s,h,g) + \int_s^t g(\xi_\kappa) \right\}.$$

Since $|\dot{\xi}_\kappa| \leq M(T)$ and $\xi_\kappa(t) = x$, hence by Arzela–Ascoli, for a subsequence still denoted by ξ_κ such that $\xi_\kappa \to \xi$ uniformly in $[s,t]$. Since $|\dot{\xi}_\kappa| \leq M(T)$ implies that $\xi \in C_{M(T)}(x,t,s)$.

Again by going to a subsequence we can assume that $\xi_\kappa \rightharpoonup \xi$ weakly in $W^{1,2}([s,t],\mathbb{R}^n)$. Hence by (4.3) and continuity of W and g it follows that

$$\begin{aligned} V(x,t) &= \lim_{\kappa\to\infty}\left\{W(\xi_\kappa(s),s) \vee \rho_+(\xi_\kappa,t,s,h,g) + \int_s^t g(\xi_\kappa)\right\} \\ &\geq W(\xi(s),s) \vee \rho_+(\xi,t,s,h,g) + \int_s^t g(\xi). \end{aligned}$$

Since $\xi \in C(x,t,s)$ and therefore by definition of V and the above inequality implies (4.8). This proves the lemma.

Lemma 4.3 (*Dynamic programming principle*). *Let u be as in* (2.8) *and* $0 \leq s < t$ *and* $x \in \mathbb{R}^n$, *then*

$$u(x,t) = \inf_{\xi\in C(x,t,s)}\left\{u(\xi(s),s) \vee \rho_+(\xi,t,s,h,g) + \int_s^t g(\xi)\right\}. \tag{4.9}$$

Proof. Let $v(x,t)$ denote the right hand side of (4.9). Let $\xi \in C(x,t)$. Then $\xi_1 = \xi|_{[0,s]} \in C(\xi(s),s)$ and $\xi_2 = \xi|_{[s,t]} \in C(x,t,s)$. Hence

$$\begin{aligned} v(x,t) &\leq u(\xi_2(s),s) \vee \rho_+(\xi_2,t,s,h,g) + \int_s^t g(\xi_2) \\ &\leq (u_0(\xi_1(0)) \vee \rho_+(\xi_1,s,h,g) + \int_0^s g(\xi_1)) \vee \rho_+(\xi_2,t,s,h,g) + \int_s^t g(\xi_2) \\ &= u_0(\xi(0)) \vee \rho_+(\xi_1,s,h,g) \vee (\rho_+(\xi_2,t,s,h,g) - \int_0^s g(\xi_1)) + \int_0^t g(\xi) \\ &= u_0(\xi(0)) \vee \rho_+(\xi,t,h,g) + \int_0^t g(\xi). \end{aligned}$$

By taking infimum over ξ, this implies that $v(x,t) \leq u(x,t)$.

Since $u_0 \in W^{1,\infty}$, by Lemma (4.2), for any $T > 0$, $|u|_T < \infty$ and hence $|v|_T < \infty$. Hence for $\epsilon > 0$, choose $\xi_2 \in C(x,t,s)$ and $\xi_1 \in C(\xi_2(s),s)$ such that

$$v(x,t) \geq u(\xi_2(s),s) \vee \rho_+(\xi_2,t,s,h,g) + \int_s^t g(\xi_2) - \epsilon,$$

$$u(\xi_2(s),s) \geq u_0(\xi_1(0)) \vee \rho_+(\xi_1,s,h,g) + \int_0^s g(\xi_1) - \epsilon.$$

Let $\xi \in C(x,t)$ be defined by $\xi|_{[0,s]} = \xi_1, \xi|_{[s,t]} = \xi_2$. Then

$$\begin{aligned} v(x,t) &\geq \left(u_0(\xi(0)) \vee \rho_+(\xi,s,h,g) + \int_0^s g(\xi) - \epsilon\right) \vee \rho_+(\xi,t,s,h,g) \\ &\qquad + \int_s^t g(\xi) - \epsilon \\ &= (u_0(\xi(0)) \vee \rho_+(\xi,s,h,g) - \epsilon) \vee \left(\rho_+(\xi,t,s,h,g) - \int_0^s g(\xi)\right) \\ &\qquad + \int_0^t g(\xi) - \epsilon \end{aligned}$$

$$
\begin{aligned}
&\geq u_0(\xi(0)) \vee \rho_+(\xi,s,h,g) \vee \left(\rho_+(\xi,t,s,h,g) - \int_0^s g(\xi)\right) + \int_0^t g(\xi) - 2\epsilon \\
&= u_0(\xi(0)) \vee \operatorname*{ess\,sup}_{\theta\in[0,s]} \left\{ h(\dot{\xi}(\theta)) - \int_0^\theta g(\xi) \right\} \\
&\quad \vee \left(\operatorname*{ess\,sup}_{\theta\in[s,t]} \left\{ h(\dot{\xi}(\theta)) - \int_s^\theta g(\xi) \right\} - \int_0^s g(\xi) \right) + \int_0^t g(\xi) - 2\epsilon \\
&= u_0(\xi(0)) \vee \operatorname*{ess\,sup}_{\theta\in[0,s]} \left\{ h(\dot{\xi}(\theta)) - \int_0^\theta g(\xi) \right\} \\
&\quad \vee \operatorname*{ess\,sup}_{\theta\in[s,t]} \left\{ h(\dot{\xi}) - \int_0^\theta g(\xi) \right\} + \int_0^t g(\xi) - 2\epsilon \\
&= u_0(\xi(0)) \vee \operatorname*{ess\,sup}_{\theta\in[0,t]} \left\{ h(\dot{\xi}(\theta)) - \int_0^\theta g(\xi) \right\} + \int_0^t g(\xi) - 2\epsilon \\
&\geq u(x,t) - 3\epsilon.
\end{aligned}
$$

Letting $\epsilon \to 0$ to obtain $v(x,t) \geq u(x,t)$ and hence $v = u$. This proves the lemma.

Lemma 4.4. *Let u be as in* (2.8). *Then for every $T > 0$, $u \in W^{1,\infty}(\mathbb{R}^n \times [0,T])$ and* $\lim_{t\to 0} u(x,t) = u_0(x)$.

Proof. Let $T > 0$, $0 < t \leq T$, $x_1, x_2 \in \mathbb{R}^n$. Since $u_0 \in W^{1,\infty}(\mathbb{R}^n)$, hence by Lemma (4.2), there exist a constant $M(T) > 0$, $\xi_1 \in C_{M(T)}(x_1,t)$ such that $|u|_T < \infty$ and $u(x_1,t) = u_0(\xi_1(0)) \vee \rho_+(\xi_1,t,h,g) + \int_0^t g(\xi_1)$. Now define $\xi_2(\theta) = \xi_1(\theta) + x_2 - x_1$, then $\xi_2 \in C(x_2,t)$ with $\dot{\xi}_1 = \dot{\xi}_2$. Let M denote the maximum of the Lipschitz constants for u_0 and g. Then

$$
\begin{aligned}
h(\dot{\xi}_2(\theta)) - \int_0^\theta g(\xi_2) &= h(\dot{\xi}_1(\theta)) - \int_0^\theta g(\xi_1) + \int_0^\theta g(\xi_1) - \int_0^\theta g(\xi_2) \\
&\leq h(\dot{\xi}_1(\theta)) - \int_0^\theta g(\xi_1) + MT|\xi_1 - \xi_2|_\infty.
\end{aligned}
$$

Hence

$$
\rho_+(\xi_2,t,h,g) \leq \rho_+(\xi_1,t,h,g) + MT|x_2 - x_1|.
$$

Therefore

$$
\begin{aligned}
u(x_2,t) &\leq u_0(\xi_2(0)) \vee \rho_+(\xi_2,t,h,g) + \int_0^t g(\xi_2) \\
&\leq (u_0(\xi_1(0)) + u_0(\xi_2(0)) - u_0(\xi_1(0))) \vee (\rho_+(\xi_1,t,h,g) \\
&\quad + MT|x_2 - x_1|) + \int_0^t g(\xi_1) + \int_0^t (g(\xi_2) - g(\xi_1)) \\
&\leq (u_0(\xi_1(0)) + M|x_2 - x_1|) \vee (\rho_+(\xi_1,t,h,g) + MT|x_2 - x_1|) \\
&\quad + \int_0^t g(\xi_1) + MT|x_2 - x_1|
\end{aligned}
$$

$$\begin{aligned}
&\le (u_0(\xi_1(0)) + M(1+T)|x_2 - x_1|) \vee (\rho_+(\xi_1, t, h, g) \\
&\quad + M(1+T)|x_2 - x_1|) + \int_0^t g(\xi_1) + MT|x_2 - x_1| \\
&\le u_0(\xi_1(0)) \vee \rho_+(\xi_1, t, h, g) + \int_0^t g(\xi_1) + 2M(1+T)|x_2 - x_1| \\
&= u(x_1, t) + 2M(1+T)|x_2 - x_1|.
\end{aligned}$$

Since x_1 and x_2 are arbitrary, the above implies that

$$|u(x_1, t) - u(x_2, t)| \le 2M(1+T)|x_2 - x_1|. \tag{4.10}$$

Let $0 \le s < t \le T$ and $x \in \mathbb{R}^n$. Since $\inf h = -\infty$ and hence there exist a $q \in \mathbb{R}^n$ such that $h(q) + T|g|_T < -|u|_T$. Let $\xi(\theta) = x + q(\theta - t) \in C(x, t, s)$. Then $|x - \xi(s)| = |q||t - s|$ and $\rho_+(\xi, t, s, h, g) \le h(q) + T|g|_T < -|u|_T \le u(\xi(s), s)$. Hence from (4.9) and (4.10)

$$\begin{aligned}
u(x,t) - u(x,s) &\le u(\xi(s), s) \vee \rho_+(\xi, t, s, h, g) + \int_s^t g(\xi) - u(x,s) \\
&\le u(\xi(s), s) - u(x,s) + |g|_T|t - s| \\
&\le 2M(1+T)|\xi(s) - x||g|_T|t - s| \\
&\le (2M(1+T)|q| + |g|_T)|t - s|.
\end{aligned} \tag{4.11}$$

Since $|u|_T < \infty$ and hence from (4.9), (4.7) and (4.10) there exist an $M(T) > 0$ such that

$$\begin{aligned}
u(x,t) &= \inf_{\xi \in C_{M(T)}(x,t,s)} \left\{ u(\xi(s), s) \vee \rho_+(\xi, t, s, h, g) + \int_s^t g(\xi) \right\} \\
&\ge \inf_{\xi \in C_{M(T)}(x,t,s)} \left\{ u(\xi(s), s) + \int_s^t g(\xi) \right\} \\
&\ge \inf_{\xi \in C_{M(T)}(x,t,s)} \{ u(\xi(s), s) - u(x,s) - M|t - s| \} + u(x,s) \\
&\ge \inf_{\xi \in C_{M(T)}(x,t,s)} \{ -2M(1+T)|\xi(s) - x| \} - M|t - s| + u(x,s) \\
&\ge -2M(1+T)M(T)|s - t| - M|t - s| + u(x,s) \\
&\ge -M_1(T)|s - t| + u(x,s),
\end{aligned}$$

where $M_1(T) = 2M(1+T)M(T) + M$. Combining this with (4.11) implies $|u(x,t) - u(x,s)| \le M_1(T)|t - s|$. By taking $s = 0$, we obtain $\lim_{t \to 0} u(x,t) = u_0(x)$ and hence the lemma.

Proof of Theorem 2.1. First we prove that u is a subsolution. Suppose not, then there exist a $(x_0, t_0) \in \mathbb{R}^n \times \mathbb{R}_+$, $\epsilon > 0$, a ball B around (x_0, t_0) and a C^1 function φ such that $\varphi(x_0, t_0) = u(x_0, t_0), u - \varphi$ has maximum at (x_0, t_0) in B and $\varphi_t + H(\varphi, D\varphi) - g \ge 4\epsilon$ at (x_0, t_0). By continuity we can choose a $\delta > 0$ such that at $(x_0, t_0), \varphi_t + H(\varphi - 2\delta, D\varphi) - g \ge 3\epsilon$. Hence from (A_3) of § 3 there exist q such that at $(x_0, t_0), \varphi_t + \langle q, D\varphi \rangle - g \ge 2\epsilon$, $h(q) \le \varphi(x_0, t_0) - 2\delta$. Now by continuity, there exist a ball $B_1 \subset B$ around (x_0, t_0) such that in B_1

$$h(q) \le \varphi - \delta, \varphi_t + \langle q, D\varphi \rangle - g \ge \epsilon. \tag{4.12}$$

Let $s_0 < t_0$ be such that the curve $\xi(\theta) = x_0 + q(\theta - t_0)$ for $\theta \in [s_0, t_0]$ is in B_1 and $\sup\{| \int_{s_0}^{\theta} g(\xi)|;\ \theta \in [s_0, t_0]\} < \delta$ from (4.9) and (4.12) we have

$$\begin{aligned}\varphi(x_0, t_0) = u(x_0, t_0) &\leq u(\xi(s_0), s_0) \vee \rho_+(\xi, t_0, s_0, h, g) + \int_{s_0}^{t_0} g(\xi)\\ &= u(\xi(s_0), s_0) \vee \left\{h(q) - \inf_{\theta\in[s_0,t_0]} \int_{s_0}^{\theta} g(\xi)\right\} + \int_{s_0}^{t_0} g(\xi)\\ &\leq u(\xi(s_0), s_0) \vee \{h(q) + \delta\} + \int_{s_0}^{t_0} g(\xi) \qquad (4.13)\\ &\leq u(\xi(s_0), s_0) \vee \varphi(\xi(s_0), s_0) + \int_{s_0}^{t_0} g(\xi)\\ &\leq \varphi(\xi(s_0), s_0) + \int_{s_0}^{t_0} g(\xi).\end{aligned}$$

Also from (4.12)

$$\begin{aligned}\varphi(x_0, t_0) - \varphi(\xi(s_0), s_0) &= \int_{s_0}^{t_0} \frac{\mathrm{d}}{\mathrm{d}\theta} \varphi(\xi(\theta), \theta)\mathrm{d}\theta\\ &= \int_{s_0}^{t_0} (\varphi_t + \langle q, D\varphi\rangle \mathrm{d}\theta) \geq \int_{s_0}^{t_0} g(\xi) + \epsilon(t_0 - s_0)\end{aligned}$$

which contradicts (4.13). This proves that u is a subsolution.

Next we prove that u is a supersolution. Suppose not, then there exists $(x_0, t_0) \in \mathbb{R}^n \times \mathbb{R}_+, \epsilon > 0$, a ball B around (x_0, t_0) and a C^1-function φ such that $u(x_0, t_0) = \varphi(x_0, t_0)$, $u - \varphi \geq 0$ in B, $\varphi_t + H(\varphi, D\varphi) - g \leq -3\epsilon$ in B. Hence from (A$_3$) of § 3, for $(x, t) \in B, q \in \mathbb{R}^n$

$$\begin{cases} (\varphi_t + \langle q, D\varphi\rangle - g)(x, t) \leq -3\epsilon,\\ \quad \text{whenever } h(q) \leq \varphi(x, t). \end{cases} \qquad (4.14)$$

From Lemma (4.4), u is continuous and $|u|_T < \infty$, for any $T > 0$. Hence from (4.9) and (4.8) for every $s < t$, there exist a $\xi_s \in C_{M(T)}(x_0, t_0, s)$ such that

$$u(x_0, t_0) = u(\xi(s), s) \vee \rho_+(\xi_s, t_0, s, h, g) + \int_s^{t_0} g(\xi_s). \qquad (4.15)$$

Since $|\dot{\xi}_s| \leq M(T)$ and hence by choosing s_0 sufficiently close to t_0, $\xi_s \in B$ for all $s \in [s_0, t_0]$.

Claim. There exist $s_1 \in [s_0, t_0)$ such that for a.e. $\theta \in [s_1, t_0]$

$$\varphi_t(\xi_{s_1}(\theta), \theta) + \langle \dot{\xi}_{s_1}(\theta), D\varphi(\xi_{s_1}(\theta), \theta)\rangle - g(\xi_{s_1}(\theta), \theta) \leq -\epsilon.$$

Suppose not, then there exist a sequence $s_m \to t_0, \theta_m \in (s_m, t_0)$ with $\xi_m = \xi_{s_m}$,

$$(\varphi_t + \langle \dot{\xi}_m, D\varphi\rangle - g)(\xi_m(\theta_m), \theta_m) > -\epsilon.$$

Let for a subsequence, $\dot{\xi}_m(\theta_m) \to q$ as $m \to \infty$. Since $\theta_m \to t_0, \xi_m(\theta_m) \to x_0$, we obtain from the above inequality

$$(\varphi_t + \langle q, D\varphi\rangle - g)(x_0, t_0) \geq -\epsilon. \qquad (4.16)$$

On the other hand from (4.15) and lower semicontinuity of h we have

$$\begin{aligned}\varphi(x_0,t_0) = u(x_0,t_0) &\geq \lim_{m\to\infty}\left\{\rho_+(\xi_m,t_0,s_m,h,g) + \int_{s_m}^{t_0} g(\xi_m)\right\}\\ &\geq \lim_{m\to\infty}\left[\left\{h(\dot{\xi}_m(\theta_m)) - \int_{s_m}^{\theta_m} g(\xi_m)\right\} + \int_{s_m}^{t_0} g(\xi_m)\right]\\ &\geq h(q).\end{aligned}$$

Hence from (4.14), $(\varphi_t + \langle q, D\varphi\rangle - g)(x_0,t_0) \leq -3\epsilon$, contradicting (4.16). This proves the claim. From the above claim we have

$$\begin{aligned}\varphi(x_0,t_0) - \varphi(\xi_{s_1}(s_1),s_1) &= \int_{s_1}^{t_0} \frac{\mathrm{d}}{\mathrm{d}\theta}\varphi(\xi_{s_1}(\theta),\theta)\mathrm{d}\theta\\ &= \int_{s_1}^{t_0} (\varphi_t + \langle \dot{\xi}_{s_1}, D\varphi\rangle(\xi_{s_1}(\theta),\theta)\mathrm{d}\theta \qquad (4.17)\\ &\leq \int_{s_1}^{t_0} g(\xi_{s_1}) - \epsilon(t_0 - s_1).\end{aligned}$$

From (4.15) we have

$$\begin{aligned}\varphi(x_0,t_0) = u(x_0,t_0) &\geq u(\xi_{s_1}(s_1),s_1) + \int_{s_1}^{t_0} g(\xi_{s_1})\\ &\geq \varphi(\xi_{s_1}(s_1),s_1) + \int_{s_1}^{t_0} g(\xi_{s_1}),\end{aligned}$$

which contradicts (4.17). This proves that u is a super solution. Furthermore from (4.8) infimum is achieved and this proves the theorem.

Theorem 2.2.

From now on we assume that H satisfies (H_3) and (H_4) of Theorem 2.2 and h be its quasi concave dual. Let ρ_- be defined as in (2.5).

Lemma 4.5. *Let W be a function on $\mathbb{R}^n \times \mathbb{R}_+$. Assume that for every $T > 0, |W|_T = \sup\{|W(x,t)| : (x,t) \in \mathbb{R}^n \times [0,T]\} < \infty$. Let $0 \leq s < t \leq T$ and $x \in \mathbb{R}^n$. Define*

$$\underline{V}(x,t) = \inf_{\xi\in C(x,t,s)}\left\{W(\xi(s),s) + \int_s^t g(\xi); W(\xi(s),s) \leq \rho_-(\xi,t,s,h,g)\right\}, \quad (4.18)$$

$$\overline{V}(x,t) = \inf_{\xi\in C(x,t,s)}\left\{W(\xi(s),s) + \int_s^t g(\xi); W(\xi(s),s) < \rho_-(\xi,t,s,h,g)\right\}, \quad (4.19)$$

then there exist a constant $M(T) > 0$ such that

$$|\underline{V}(x,t)| \vee |\overline{V}(x,t)| \leq |W|_T + T|g|_\infty, \quad (4.20)$$

$$\underline{V}(x,t) = \inf_{\xi\in C_{M(T)}(x,t,s)}\left\{W(\xi(s),s) + \int_s^t g(\xi); W(\xi(s),s) \leq \rho_-(\xi,t,s,h,g)\right\}, \quad (4.21)$$

$$\overline{V}(x,t) = \inf_{\xi\in C_{M(T)}(x,t,s)}\left\{W(\xi(s),s) + \int_s^t g(\xi); W(\xi(s),s) < \rho_-(\xi,t,s,h,g)\right\}. \quad (4.22)$$

Furthermore if W is lower semicontinuous function, then there exist a $\xi \in C_{M(T)}(x,t,s)$ such that

$$W(\xi(s),s) \le \rho_-(\xi,t,s,h,g), \quad \underline{V}(x,t) = W(\xi(s),s) + \int_s^t g(\xi). \tag{4.23}$$

If W is continuous, then there exist $\{\eta_\kappa\} \subset C_{M(T)}(x,t,s), \eta \in C_{M(T)}(x,t,s)$ such that $\eta_\kappa \to \eta$ in C^0 and

$$\overline{V}(x,t) = W(\eta(s),s) + \int_s^t g(\eta), \tag{4.24}$$

$$W(\eta_\kappa(s),s) < \rho_-(\eta_\kappa,t,s,h,g). \tag{4.25}$$

Proof. Since $-(|W|_T + T|g|_\infty) \le W(\xi(s),s) + \int_s^t g(\xi) \le (|W|_T + T|g|_\infty)$ and hence (4.20) follows. Since $\lim_{|p|\to\infty} h(p) = -\infty$, there exist $M(T) > 0$ such that if $|p| > M(T)$, then $h(p) < -2(|W|_T + T|g|_\infty)$. Let $\xi \in C(x,t,s)$ such that $|\dot{\xi}|_\infty > M(T)$. Then for θ in a set of positive measure in $[s,t]$

$$h(\dot{\xi}(\theta)) - \int_s^\theta g(\xi) \le -2(|W|_T + T|g|_\infty) + T|g|_\infty \le -2|W|_\infty,$$

and hence $\rho_-(\xi,t,s,h,g) < W(\xi(s),s)$. This proves (4.21) and (4.22).

Let $\{\xi_\kappa\}$ be a minimizing sequence in (4.18). By going to a subsequence we can assume that $\xi_\kappa \to \xi$ in C^0 and $\xi_\kappa \to \xi$ weakly in $W^{1,2}([s,t],\mathbb{R}^n)$. Hence from $(\mathrm{A}_4), (\mathrm{A}_5)$ of § 3, from (4.4) and by lower semicontinuity of W we have

$$W(\xi(s),s) \le \underline{\lim}_{\kappa\to\infty} W(\xi_\kappa(s),s) \le \lim_{\kappa\to\infty} \rho_-(\xi_\kappa,t,s,h,g) \le \rho_-(\xi,t,s,h,g).$$

Hence

$$W(\xi(s),s) + \int_0^t g(\xi) \ge \underline{V}(x,t) = \lim\left\{W(\xi_\kappa(s),s) + \int_0^t g(\xi_\kappa)\right\}$$
$$\ge W(\xi(s),s) + \int_0^t g(\xi).$$

This proves (4.23). Any minimizing sequence $\{\eta_\kappa\} \subset C_{M(T)}(x,t,s)$ of $\overline{V}$, we can extract a subsequence and still denote it by $\{\eta_\kappa\}$ converging strongly to η in C^0-topology. Now from continuity of W, (4.24) and (4.25) follow.

Lemma 4.6. Let $g(x,t) = g_1(x,t) + g_2(t)$. Assume that $tg_1(x,t)$ is non-increasing in $t, g_2(t) \le 0$ and $H(u,p) > 0$ for all $u \in \mathbb{R}, |p| = 1$. Then $\underline{u}^* = \overline{u}, \quad \overline{u}_* = \underline{u}$.

Proof. The proof is divided into three steps.

Step 1. Let $\alpha > 1$ and $\{q_k\}$ be a bounded sequence. Then $\underline{\lim}_{k\to\infty} h(\alpha q_k) < \underline{\lim}_{k\to\infty} h(q_k)$.

Suppose not, then let for a subsequence still denoted by $\{q_k\}$ such that

$$q_k \to q_0, \quad \lim_{k\to\infty} h(\alpha q_k) = \lim_{k\to\infty} h(q_k) = \eta.$$

Choose $|p_k| = |\tilde{p}_k| = 1$ such that for all $|p| = 1$

(i) $H(h(q_k),p_k) = \langle q_k,p_k\rangle, \quad H(h(q_k),p) \ge \langle q_k,p\rangle$

(ii) $H(h(\alpha q_k),\tilde{p}_k) = \alpha\langle q_k,\tilde{p}_k\rangle, \quad H(h(\alpha q_k),p) \ge \alpha\langle q_k,p\rangle$.

Again going to a subsequence, one can assume that $p_k \to p_0$ $\tilde{p}_k \to \tilde{p}_0$ as $k \to \infty$. Then by continuity of H, $\langle q_0, p_0\rangle = H(\eta, p_0) = \lim_{k\to\infty} H(h(\alpha q_k), p_0) \geq \alpha\langle q_0, p_0\rangle$. Since $H(\eta, p_0) > 0$, it follows that $\alpha \leq 1$ which is a contradiction. This proves step 1.

Step 2. Let $t_1 > t$, then $\underline{u}(x,t) \geq \overline{u}(x, t_1)$. Let $\xi \in C_M(x,t)$ such that

$$\underline{u}(x,t) = u_0(\xi(0)) + \int_0^t g(\xi); \quad u_0(\xi(0)) \leq \operatorname*{ess\,inf}_{\theta\in[0,t]} \left\{ h(\dot{\xi}(\theta)) - \int_0^\theta g(\xi) \right\}.$$

Let $\xi_1(\theta) = \xi(\frac{t\theta}{t_1})$ for $\theta \in [0, t_1]$. Then $\xi_1 \in C(x, t_1)$ and $\xi_1(0) = \xi(0)$. Choose a sequence $\theta_k \in [0, t_1]$ and from step (1) to obtain

$$\begin{aligned}
\operatorname*{ess\,inf}_{\theta\in[0,t_1]} \left\{ h(\dot{\xi}_1(\theta)) - \int_0^{(t/t_1)\theta} g(\xi) \right\} &= \lim_{k\to\infty} \left\{ h(\dot{\xi}_1(\theta_k)) - \int_0^{(t/t_1)\theta_k} g(\xi) \right\} \\
&> \lim_{k\to\infty} \left\{ h\Big(\frac{t_1}{t}\dot{\xi}_1(\theta_k)\Big) - \int_0^{(t/t_1)\theta_k} g(\xi) \right\} \\
&= \lim_{k\to\infty} \left\{ h\Big(\dot{\xi}\Big(\frac{t}{t_1}\theta_k\Big)\Big) - \int_0^{(t/t_1)\theta_k} g(\xi) \right\} \\
&\geq \operatorname*{ess\,inf}_{\theta\in[0,t]} \left\{ h(\dot{\xi}(\theta)) - \int_0^\theta g(\xi) \right\}.
\end{aligned}$$

Hence from tg_1 non-increasing in t and $g_2(t) \leq 0$ we have that

$$\begin{aligned}
u_0(\xi_1(0)) &\leq \operatorname*{ess\,inf}_{\theta\in[0,t]} \left\{ h(\dot{\xi}(\theta)) - \int_0^\theta g(\xi) \right\} \\
&< \operatorname*{ess\,inf}_{\theta\in[0,t_1]} \left\{ h(\dot{\xi}_1(\theta)) - \int_0^{(t/t_1)\theta} g(\xi_1) \right\} \\
&= \operatorname*{ess\,inf}_{\theta\in[0,t_1]} \left\{ h(\dot{\xi}_1(\theta)) - \int_0^\theta g_1(\xi_1) + \int_0^\theta g_1(\xi_1) - \int_0^{(t/t_1)\theta} g_1(\xi) \right. \\
&\qquad \left. - \int_0^{(t/t_1)\theta} g_2(s)ds + \int_0^\theta g_2(s)ds - \int_0^\theta g_2(s)ds \right\} \\
&\leq \operatorname*{ess\,inf}_{\theta\in[0,t_1]} \left\{ h(\dot{\xi}_1(\theta)) - \int_0^\theta g_1(\xi_1) - \int_0^\theta g_2(s)ds \right\} \\
&= \operatorname*{ess\,inf}_{\theta\in[0,t_1]} \left\{ h(\dot{\xi}_1(\theta)) - \int_0^\theta g(\xi_1) \right\}.
\end{aligned}$$

Since $\int_0^\theta g_1(\xi_1) - \int_0^{(t/t_1)\theta} g_1(\xi) = \int_0^{(t/t_1)\theta} (\frac{t_1}{t} g_1(\xi(s), \frac{t_1}{t}s) - g_1(\xi(s), s))ds \leq 0$ and $g_2 \leq 0$, this implies that

$$\begin{aligned}
\underline{u}(x,t) &= u_0(\xi(0)) + \int_0^t g(\xi) \\
&= u_0(\xi_1(0)) + \int_0^{t_1} g(\xi_1) + \int_0^t g(\xi) - \int_0^{t_1} g(\xi_1) \\
&\geq \overline{u}(x, t_1).
\end{aligned}$$

Step 3. Let $B_r(x,t)$ be a ball centered at (x,t) with radius r. Let $t > 0$ and let $t_k < t$ and $t_k \to t$. Then from step 2,

$$\begin{aligned}\underline{u}^*(x,t) &= \lim_{r\to 0}\sup_{B_r}\underline{u}(z)\\ &\geq \lim_{k\to\infty}\underline{u}(x,t_k) \geq \overline{u}(x,t).\end{aligned}$$

On the other hand $\underline{u} \leq \overline{u}$ and hence $\underline{u}^*(x,t) \leq \overline{u}(x,t)$, implies that $\underline{u}^*(x,t) = \overline{u}(x,t)$. Similarly $\overline{u}_* = \underline{u}$. This proves the Lemma.

In order to prove, the representation formula for a solution in the sense of viscosity, one has to establish a dynamic programming principle. This has been carried out for standard control problems and differential games in [2, 8] and [10]. We will next provide a proof of this fact for our problem.

Lemma 4.7 (*Dynamic programming principle*). *For every* $T > 0$, *there exist* $M(T) > 0$ *such that* $|\underline{u}|_T < \infty, \underline{u}$ *is lower semicontinuous and for* $0 \leq s < t \leq T, x \in \mathbb{R}^n$,

$$\underline{u}(x,t) = \inf_{C_{M(t)}(x,t,s)}\left\{\underline{u}(\xi(s),s) + \int_s^t g(\xi) : \underline{u}(\xi(s),s) \leq \rho_-(\xi,t,s,h,g)\right\}, \tag{4.26}$$

$$\underline{u}^*(x,t) \leq \inf_{C_{M(T)}(x,t,s)}\left\{\underline{u}^*(\xi(s),s) + \int_0^t g(\xi); \underline{u}^*(\xi(s),s) < \rho_-(\xi,t,s,h,g)\right\}. \tag{4.27}$$

Proof. Since $u_0 \in W^{1,\infty}(\mathbb{R}^n)$ and hence by taking $s = 0$ in (4.18), $|\underline{u}|_T < \infty$ follows from (4.20). Let $(x_m,t_m) \to (x,t)$ as $m \to \infty$. Since u_0 is continuous, from (4.23), for each m, there exist a $\xi_m \in C_{M(T)}(x_m,t_m)$ such that $\underline{u}(x_m,t_m) = u_0(\xi_m(0)) + \int_0^{t_m} g(\xi_m)$ and $u_0(\xi_m(0)) \leq \rho_-(\xi_m,t_m,h,g)$. From Arzela–Ascoli we can extract a subsequence still denoted by ξ_m such that $\xi_m \to \xi \in C_{M(T)}(x,t)$. Now from (4.4) we have

$$u_0(\xi(0)) \leq \overline{\lim_{m\to\infty}}\,\rho_-(\xi_m,t_m,h,g) \leq \rho_-(\xi,t,h,g),$$

therefore,

$$\underline{\lim_{m\to\infty}}\,\underline{u}(x_m,t_m) = u_0(\xi(0)) + \int_0^t g(\xi) \geq \underline{u}(x,t).$$

This proves $\underline{u}$ is lower semicontinuous. Let

$$v_1(x,t) = \inf_{C(x,t,s)}\left\{\underline{u}(\xi(s),s) + \int_s^t g(\xi); \underline{u}(\xi(s),s) \leq \rho_-(\xi,t,s,h,g)\right\}. \tag{4.28}$$

Since $|\underline{u}|_T < \infty$ and $\underline{u}$ is lower semicontinuous, hence from (4.23), there exist a $\xi_2 \in C(x,t,s)$ such that $\underline{u}(\xi_2(s),s) \leq \rho_-(\xi_2,t,s,h,g)$ and $v_1(x,t) = \underline{u}(\xi_2(s),s) + \int_s^t g(\xi_2)$. Choose a $\xi_1 \in C(\xi_2(s),s)$ such that $\underline{u}(\xi_2(s),s) = u_0(\xi_1(0)) + \int_0^s g(\xi_1)$, $u_0(\xi_1(0)) \leq \rho_-(\xi_1,s,h,g)$. Let $\eta \in C(x,t)$ defined by $\eta(\theta) = \xi_1(\theta)$ for $\theta \in [0,s]$ and $\eta(\theta) = \xi_2(\theta)$ for $\theta \in [s,t]$. Then we have

$$u_0(\eta(0)) = \underline{u}(\xi_1(s),s) - \int_0^s g(\xi_1)$$

$$
\begin{aligned}
&\le \rho_-(\xi_2, t, s, h, g) - \int_0^s g(\xi_1)\\
&= \operatorname*{ess\,inf}_{\theta\in[s,t]} \left\{ h(\dot\eta(\theta)) - \int_0^\theta g(\eta) \right\}.
\end{aligned}
$$

Since $u_0(\eta(0)) \le \rho_-(\eta, s, h, g)$, it follows that $u_0(\eta(0)) \le \rho_-(\eta, t, h, g)$. Therefore we have

$$
\begin{aligned}
v_1(x,t) &= \underline{u}(\eta(s), s) + \int_s^t g(\eta)\\
&= u_0(\eta(0)) + \int_0^t g(\eta) \ge \underline{u}(x,t).
\end{aligned}
\tag{4.29}
$$

Let $\xi \in C(x,t)$ be such that $\underline{u}(x,t) = u_0(\xi(0)) + \int_0^t g(\xi)$ and $u_0(\xi(0)) \le \rho_-(\xi, t, h, g) = \inf\{\operatorname{ess\,inf}_{\theta\in[0,s]}\{h(\dot\xi(\theta)) - \int_0^\theta g(\xi)\}, \operatorname{ess\,inf}_{\theta\in[s,t]}\{h(\xi(\theta)) - \int_0^\theta g(\xi)\}\}$. Hence

$$
\begin{aligned}
\underline{u}(\xi(s), s) &\le u_0(\xi(0)) + \int_0^s g(\xi)\\
&\le \operatorname*{ess\,inf}_{\theta\in[s,t]} \left\{ h(\dot\xi(\theta) - \int_0^\theta g(\xi) + \int_0^s g(\xi) \right\}\\
&= \rho_-(\xi, t, s, h, g).
\end{aligned}
$$

This implies that $\underline{u}(x,t) = u_0(\xi(0)) + \int_0^s g(\xi) + \int_s^t g(\xi) \ge \underline{u}(\xi(s), s) + \int_s^t g(\xi) \ge v_1(x,t)$. Therefore from (4.29) $\underline{u}(x,t) = v_1(x,t)$ and since $|\underline{u}|_T < \infty$ and hence from (4.21), (4.26) follows.

Let

$$
v_2(x,t) = \inf_{\xi\in C(x,t,s)} \left\{ \underline{u}^*(\xi(s), s) + \int_s^t g(\xi); \underline{u}^*(\xi(s), s) < \rho_-(\xi, t, s, h, g) \right\}.
\tag{4.30}
$$

Choose a sequence $(x_\kappa, t_\kappa) \to (x,t)$ such that $\underline{u}^*(x,t) = \lim_{\kappa\to\infty} \underline{u}(x_\kappa, t_\kappa)$. For $\xi \in C(x,t,s)$, let $s_\kappa = s + t_\kappa - t$ and define $\xi_\kappa \in C(x_\kappa, t_\kappa, s_\kappa)$ by $\xi_\kappa(\theta) = \xi(\theta - t_\kappa + t) + x_\kappa - x$. Then by change of variables $\alpha = \theta - t_\kappa + t$, we obtain

$$
\operatorname*{ess\,inf}_{\theta\in[s_\kappa,t_\kappa]} \left\{ h(\dot\xi_\kappa(\theta)) - \int_{s_\kappa}^\theta g(\xi_\kappa) \right\} = \operatorname*{ess\,inf}_{\alpha\in[s,t]} \left\{ h(\dot\xi(\alpha)) - \int_s^\alpha g(\xi) \right\} + \beta_\kappa,
\tag{4.31}
$$

where $\beta_\kappa = O(\sup_{\alpha\in[s,t]}(\int_s^\alpha g(\xi) - \int_{s_\kappa}^{\alpha-t+t_\kappa} g(\xi_\kappa))) \to 0$ as $\kappa \to \infty$.

Let $\xi \in C(x,t,s)$ such that $\underline{u}^*(\xi(s), s) < \rho_-(\xi, t, s, h, g)$. Hence from (4.29) and upper semicontinuity of $\underline{u}^*$ we can find a $\kappa_0 > 0$ such that for $\kappa \ge \kappa_0$, $\underline{u}^*(\xi_\kappa(s_\kappa), s_\kappa) < \rho_-(\xi_\kappa, t_\kappa, s_\kappa, h, g)$ and hence $\underline{u}(\xi_\kappa(s_\kappa), s_\kappa) \le \underline{u}^*(\xi_\kappa(s_\kappa), s_\kappa) < \rho_-(\xi_\kappa, t_\kappa, s_\kappa, h, g)$. Therefore from (4.28) we have

$$
\begin{aligned}
\underline{u}^*(x,t) &= \lim_{\kappa\to\infty} \underline{u}(x_\kappa, t_\kappa) \le \lim_{\kappa\to\infty} v_1(x_\kappa, t_\kappa)\\
&\le \lim_{\kappa\to\infty} \left\{ \underline{u}(\xi_\kappa(s_\kappa), s_\kappa) + \int_{s_\kappa}^{t_\kappa} g(\xi_\kappa) \right\}
\end{aligned}
$$

$$\leq \lim_{\kappa\to\infty}\left\{\underline{u}^*(\xi_\kappa(s_\kappa),s_\kappa)+\int_{s_\kappa}^{t_\kappa} g(\xi_\kappa)\right\}$$

$$\leq \underline{u}^*(\xi(s),s)+\int_s^t g(\xi).$$

Since it is true for all ξ and hence $\underline{u}^*(x,t) \leq v_2(x,t)$. Combining this with $|\underline{u}^*|_T < \infty$ and (4.22) we obtain (4.27). This proves the Lemma.

Lemma 4.8 (*Dynamic programming principle*). *For every $T > 0, |\overline{u}|_T < \infty, \overline{u}$ is upper semicontinuous and there exist $M(T) > 0$ such that for $0 \leq s < t \leq T, x \in \mathbb{R}^n$,*

$$\overline{u}(x,t)=\inf_{C_{M(T)}(x,t,s)}\left\{\overline{u}(\xi(s),s)+\int_s^t g(\xi);\ \overline{u}(\xi(s),s)< \rho_-(\xi,t,s,h,g)\right\}, \tag{4.32}$$

$$\overline{u}_*(x,t) \geq \inf_{C_{M(T)}(x,t,s)}\left\{\overline{u}_*(\xi(s),s)+\int_s^t g(\xi);\ \overline{u}_*(\xi(s),s)\leq \rho_-(\xi,t,s,h,g)\right\}. \tag{4.33}$$

Proof. Since $u_0 \in W^{1,\infty}(\mathbb{R}^n)$, by taking $s=0$ in (4.18), $|\overline{u}|_T < \infty$ follows from (4.20). Let $(x_m,t_m) \to (x,t)$ as $m \to \infty$. Since u_0 is continuous, by (4.24) and (4.25) there exist $\eta, \eta_\kappa \in C_{M(T)}(x,t)$ such that $\eta_\kappa \to \eta$ uniformly and $\overline{u}(x,t) = u_0(\eta(0)) + \int_0^t g(\eta)$ and $u_0(\eta_\kappa(0)) < \rho_-(\eta_\kappa,t,h,g)$. Now for each κ, define $\eta_{m_\kappa} \in C(x_m,t_m)$ as follows:

$$\eta_{m_\kappa}(\theta)=\begin{cases}\eta_\kappa(\theta-t_m+t)+x_m-x & \text{if} \quad \theta\in[0\vee(t_m-t),t_m],\\ \eta_\kappa(0)+x_m-x & \text{if} \quad \theta\in[0,0\vee(t_m-t)].\end{cases} \tag{4.34}$$

Clearly $\eta_{m_\kappa} \to \eta_\kappa$ uniformly and by change of variables it follows that

$$\rho_-(\eta_{m_\kappa},t_m,h,g)=\rho_-(\eta_\kappa,t,h,g)+o(1), \tag{4.35}$$

where $o(1) \to 0$ as $m \to \infty$. Since u_0 is continuous, from (4.35) we can find a $m(\kappa) > 0$ such that for $m > m(\kappa), u_0(\eta_{m_\kappa}(0)) < \rho_-(\eta_{m_\kappa},t_m,h,g)$. This implies that $\overline{u}(x_m,t_m) \leq u_0(\eta_{m_\kappa}(0)) + \int_0^{t_m} g(\eta_{m_\kappa})$. Now letting $m \to \infty, \kappa \to \infty$, we conclude that

$$\overline{\lim_{m\to\infty}}\,\overline{u}(x_m,t_m)\leq \lim_{\kappa\to\infty}\left\{u_0(\eta_\kappa(0))+\int_0^t g(\eta_\kappa)\right\}=\overline{u}(x,t).$$

This proves $\overline{u}$ is upper semicontinuous. Define

$$v_1(x,t)=\inf_{\xi\in C(x,t,s)}\left\{\overline{u}(\xi(s),s)+\int_s^t g(\xi)\ :\ \overline{u}(\xi(s),s)<\rho_-(\xi,t,s,h,g)\right\},$$

then

$$v_1(x,t)=\inf_{\xi\in C(x,t,s)}\left[\inf_{\eta\in C(\xi(s),s)}\left\{u_0(\eta(0))+\int_0^s g(\eta)+\int_s^t g(\xi);\ u_0(\eta(0))\right.\right.$$
$$\left.< \rho_-(\eta,s,h,g)\right\};\quad \inf_{\eta\in C(\xi(s),s)}\left\{u_0(\eta(0))+\int_0^s g(\eta) : u_0(\eta(0))\right.$$
$$\left.\left.< \rho_-(\eta,s,h,g)\right\}<\rho_-(\xi,t,s,h,g)\right]$$

$$\begin{aligned}&\le \inf_{\lambda\in C(x,t)}\left\{u_0(\lambda(0)) + \int_0^t g(\lambda); u_0(\lambda(0)) < \rho_-(\lambda,s,h,g), u_0(\lambda(0))\right.\\ &\qquad\left. + \int_0^s g(\lambda) < \rho_-(\lambda,t,s,h,g)\right\}\\ &= \inf_{\lambda\in C(x,t)}\left\{u_0(\lambda(0)) + \int_0^t g(\lambda) : u_0(\lambda(0)) < \rho_-(\lambda,t,h,g)\right\}\\ &= \overline{u}(x,t).\end{aligned} \tag{4.36}$$

Since $|\overline{u}|_T < \infty$, hence from (4.22), there exist a $M_1(T) > 0$ such that

$$v_1(x,t) = \inf_{\xi\in C_{M_1(T)}(x,t,s)}\left\{\overline{u}(\xi(s),s) + \int_s^t g(\xi);\ \overline{u}(\xi(s),s) < \rho_-(\xi,t,s,h,g)\right\}. \tag{4.37}$$

Let $\varepsilon > 0, r > 0$ and $\xi \in C(x,t,s)$ such that $v_1(x,t) \ge \overline{u}(\xi(s),s) + \int_s^t g(\xi) - \varepsilon$ and $\overline{u}(\xi(s),s) + r < \rho_-(\xi,t,s,h,g)$. From (4.24), (4.25) and (4.4) there exist an $\eta \in C(\xi(s),s)$ such that $\overline{u}(\xi(s),s) > u_0(\eta(0)) + \int_0^s g(\eta) - r,\quad u_0(\eta(0)) < \rho_-(\eta,s,h,g)$. Let $\lambda \in C(x,t)$ be defined by $\lambda|_{[0,s]} = \eta, \lambda|_{[s,t]} = \xi$. Then $u_0(\lambda(0)) = u_0(\eta(0))$ and

$$\begin{aligned}u_0(\lambda(0)) &= \overline{u}(\xi(s),s) + r - \int_0^s g(\eta)\\ &< \rho_-(\xi,t,s,h,g) - \int_0^s g(\eta)\\ &= \operatorname*{ess\,inf}_{\theta\in[s,t]}\left\{h(\dot\lambda(\theta)) - \int_0^\theta g(\lambda)\right\}.\end{aligned}$$

Since $u_0(\lambda(0)) = u_0(\eta(0)) < \rho_-(\eta,s,h,g)$ and hence combining this with the above inequality implies that $u_0(\lambda(0)) < \inf\{\operatorname{ess\,inf}_{\theta\in[s,t]}\{h(\dot\lambda(\theta)) - \int_0^\theta g(\lambda)\}$, $\operatorname{ess\,inf}_{\theta\in[0,s]}$ $\{h(\dot\lambda(\theta)) - \int_0^\theta g(\lambda)\}\} = \rho_-(\lambda,\ t,\ h,\ g)$. Therefore $v_1(x,\ t) \ge u_0(\lambda(0)) + \int_0^t g(\lambda) - \epsilon \ge\ \overline{u}(x,\ t) - \epsilon$. Since ϵ is arbitrary, we obtain $v_1(x,t) \ge \overline{u}(x,t)$. This with (4.36), (4.37) implies (4.32)

$$v_2(x,t) = \inf_{\xi\in C(x,t,s)}\left\{\overline{u}_*(\xi(s),s) + \int_s^t g(\xi); \overline{u}_*(\xi(s),s) \le \rho_-(\xi,t,s,h,g)\right\}. \tag{4.38}$$

Let $\lim_{\kappa\to\infty}(x_\kappa,t_\kappa) = (x,t), \lim_{\kappa\to\infty}\overline{u}(x_\kappa,t_\kappa) = \overline{u}_*(x,t)$. Let $\epsilon > 0$. Then from (4.36), (4.37) and (4.38) we can choose a $\kappa(\epsilon) > 0$ such that for every $\kappa > \kappa(\epsilon)$, there exist a $\xi_\kappa \in C_{M_1(T)}(x_\kappa,t_\kappa,s_\kappa)$ such that

$$\overline{u}_*(x,t) \ge \overline{u}(x_\kappa,t_\kappa) - \frac{\epsilon}{2}, \tag{4.39}$$

$$\overline{u}(x_\kappa,t_\kappa) > \overline{u}(\xi_\kappa(s),s) + \int_s^{t_\kappa} g(\xi_\kappa) - \frac{\epsilon}{2}, \tag{4.40}$$

$$\overline{u}(\xi_\kappa(s),s) < \rho_-(\xi_\kappa,t_\kappa,s,h,g). \tag{4.41}$$

Extract a subsequence still denoted by ξ_κ converging to ξ uniformly. Then from (4.41), (4.4), (4.39) and (4.40)

$$\overline{u}_*(\xi(s),s) \le \varliminf_{\kappa\to\infty} \overline{u}_*(\xi_\kappa(s),s) \le \lim_{\kappa\to\infty} \overline{u}_*(\xi_\kappa(s),s) \le \rho_-(\xi,t,s,h,g), \tag{4.42}$$

$$\begin{aligned}\overline{u}_*(x,t) &\geq \lim_{\kappa\to\infty} \overline{u}(x_\kappa,t_\kappa) - \frac{\epsilon}{2}\\ &\geq \underline{\lim}_{\kappa\to\infty} \left\{\overline{u}(\xi_\kappa(s),s) + \int_s^{t_\kappa} g(\xi_\kappa)\right\} - \epsilon\\ &\geq \underline{\lim}_{\kappa\to\infty} \left\{\overline{u}_*(\xi_\kappa(s),s) + \int_s^{t_\kappa} g(\xi_\kappa)\right\} - \epsilon\\ &= \overline{u}_*(\xi(s),s) + \int_s^t g(\xi) - \epsilon\\ &\geq v_2(x,t) - \epsilon,\end{aligned}$$

since (4.42) holds. Now letting $\epsilon \to 0$ to conclude that $\overline{u}_*(x,t) \geq v_2(x,t)$. Since $|\overline{u}_*|_T < \infty$ and hence from (4.21) there exist an $M(T) > 0$ such that (4.33) holds. This proves the lemma.

Proof of Theorem 2.2. Let $T > 0$ and $0 < t \leq T$ and $x \in \mathbb{R}^n$. From (4.23) and (4.24) there exist $M(T) > 0$ such that $\xi_t, \eta_t \in C_{M(T)}(x,t)$ and

$$\underline{u}(x,t) = u_0(\xi_t(0)) + \int_0^t g(\xi_t),$$

$$\overline{u}(x,t) = u_0(\eta_t(0)) + \int_0^t g(\eta_t).$$

Since $|(x,x) - (\xi_t(\theta), \eta_t(\theta))| = |\int_\theta^t (\dot{\xi}_t(\lambda), \dot{\eta}_t(\lambda))\mathrm{d}\lambda| \leq M(T)|t-\theta|$ and hence $(\xi_t, \eta_t) \to (x,x)$ as $t \to 0$. This implies that $\lim_{t\to 0}(\underline{u}(x,t), \overline{u}(x,t)) = (u_0(x), u_0(x))$.

Suppose $\underline{u}$ is not a sub solution. Then there exist an $(x_0,t_0) \in \mathbb{R}^n \times \mathbb{R}_+, \epsilon > 0, B$ a ball with centre (x_0,t_0) and a $\varphi \in C^1(\mathbb{R}^n \times \mathbb{R}_+)$ such that $\underline{u}^*(x_0,t_0) = \varphi(x_0,t_0), \underline{u}^* - \varphi \leq 0$ in $B, \varphi_t + H(\varphi, D\varphi) - g \geq 4\epsilon$ at (x_0,t_0). By continuity we can choose a $\delta > 0$ such that $\varphi_t + H(\varphi+\delta, D\varphi) - g \geq 3\epsilon$ at (x_0,t_0). Therefore from (A_6) of § 3, there exist q with $\varphi(x_0,t_0) + \delta \leq h(q)$ and $\varphi_t + < q, D\varphi > -g \geq 2\epsilon$ at (x_0,t_0). By continuity, we can find a ball $B_1 \subset B$ around (x_0,t_0) such that for $(x,t) \in B_1$,

$$\underline{u}^*(x,t) \leq \varphi(x,t) < h(q) - \frac{\delta}{2}, \tag{4.43}$$

$$\varphi_t + \langle q, D\varphi\rangle - g \geq \epsilon. \tag{4.44}$$

Let $\xi(\theta) = x_0 + q(\theta - t_0)$ and choose a $s_0 < t_0$ such that for $\theta \in [s_0,t_0]$, $\sup_{\theta\in[s_0,t_0]} |\int_{s_0}^\theta g(\xi)| < \frac{\delta}{2}$ and $(\xi(\theta),\theta) \in B_1$. Then from (4.43), $\underline{u}^*(\xi(s_0),s_0) < h(q) - \frac{\delta}{2} \leq \inf_{\theta\in[s_0,t_0]} \{h(\dot{\xi}(\theta)) - \int_{s_0}^\theta g(\xi)\} = \rho_-(\xi, t_0, s_0, h, g)$. Hence from (4.26)

$$\begin{aligned}\varphi(x_0,t_0) = \underline{u}^*(x_0,t_0) &\leq \underline{u}^*(\xi(s_0),s_0) + \int_{s_0}^{t_0} g(\xi)\\ &\leq \varphi(\xi(s_0),s_0) + \int_{s_0}^{t_0} g(\xi).\end{aligned} \tag{4.45}$$

From (4.44) we have

$$\varphi(x_0,t_0) - \varphi(\xi(s_0),s_0) = \int_{s_0}^{t_0} \frac{\mathrm{d}}{\mathrm{d}\theta}(\varphi(\xi(\theta),\theta))\mathrm{d}\theta$$

$$= \int_{s_0}^{t_0} (\varphi_t + \langle q, D\varphi\rangle)(\xi(\theta), \theta)\mathrm{d}\theta$$
$$\geq \int_{s_0}^{t_0} g(\xi) + \epsilon(t_0 - s_0),$$

which contradicts (4.45). This proves $\underline{u}$ is a sub solution.

Next we will show that $\underline{u}$ is a super solution. Suppose not, since $\underline{u}$ is a lower semi-continuous function, hence there exist a $(x_0, t_0) \in \mathbb{R}^n \times \mathbb{R}_+, \epsilon > 0$ a ball B centered at (x_0, t_0) and a $\varphi \in C^1(\mathbb{R}^n \times \mathbb{R}_+)$ such that $\underline{u}(x_0, t_0) = \varphi(x_0, t_0), \underline{u} - \varphi \geq 0$ in B, $\varphi_t + H(\varphi, D\varphi) - g \leq -4\epsilon$ at (x_0, t_0). Hence by continuity of H and (A_6) of § 3, we can find a ball $B_1 \subset B$ centered at (x_0, t_0) such that $\underline{u} \geq \varphi$ in B_1 and whenever $q \in \mathbb{R}^n, (x, t) \in B_1$ with $\varphi(x, t) \leq h(q)$, then at (x, t)

$$\varphi_t + \langle q, D\varphi\rangle - g \leq -2\epsilon. \tag{4.46}$$

For every $s \leq t_0$, from (4.23) and (4.26) choose a $\xi_s \in C_{M(T_0)}(x_0, t_0, s)$ such that $\underline{u}(\xi_s(s), s) \leq \rho_-(\xi_s, t_0, s, h, g)$ and $\underline{u}(x_0, t_0) = \underline{u}(\xi_s(s), s) + \int_s^{t_0} g(\xi_s)$. Now $|\xi_s(\theta) - x_0| \leq M(T_0)|t_0 - \theta|$ and hence we can find a $s_0 < t_0$ such that for any $s \in [s_0, t_0], (\xi_s(\theta), \theta) \in B_1$ for all $\theta \in (s_0, t_0]$. Therefore for $s \in [s_0, t_0]$

$$\begin{aligned} \varphi(x_0, t_0) = \underline{u}(x_0, t_0) &= \underline{u}(\xi_s(s), s) + \int_s^{t_0} g(\xi_s) \\ &\geq \varphi(\xi_s(s), s) + \int_s^{t_0} g(\xi_s) \end{aligned} \tag{4.47}$$

$$\varphi(\xi_s(s), s) \leq \underline{u}(\xi_s(s), s) \leq \rho_-(\xi_s, t_0, s, h, g). \tag{4.48}$$

Claim. There exist $s_1 \in [s_0, t_0)$ such that for almost every $\theta \in [s_1, t_0]$

$$\varphi_t(\xi_{s_1}(\theta), \theta) + \langle \dot{\xi}_{s_1}(\theta), D\varphi(\xi_{s_1}(\theta), \theta)\rangle - g(\xi_{s_1}(\theta), \theta) \leq -\epsilon. \tag{4.49}$$

Suppose not, then from (4.48) we can find a sequence $s_m \to t_0, \theta_m \in [s_m, t_0]$, $\xi_m = \xi_{s_m}$ such that

$$\varphi_t(\xi_m(\theta_m), \theta_m) + \langle \dot{\xi}_m(\theta_m), D\varphi(\xi_m(\theta_m), \theta_m)\rangle - g(\xi_m(\theta_m), \theta_m) \geq -\epsilon \tag{4.50}$$

$$\varphi(\xi_m(\theta_m), \theta_m) \leq h(\dot{\xi}_m(\theta_m))) - \int_{s_m}^{\theta_m} g(\xi_m). \tag{4.51}$$

Since $|\dot{\xi}_m(\theta_m)| \leq M(t_0)$, hence for a subsequence still denoted by ξ_m, let $q = \lim_{m\to\infty} \dot{\xi}_m(\theta_m)$. Now letting $m \to \infty$ in (4.50) and (4.51) and using upper semicontinuity of h to obtain

$$\varphi_t(x_0, t_0) + \langle q, D\varphi(x_0, t_0)\rangle - g(x_0, t_0) \geq -\epsilon$$
$$\varphi(x_0, t_0) \leq \overline{\lim_{m\to\infty}} \left\{ h(\dot{\xi}_m(\theta_m)) - \int_{s_m}^{\theta_m} g(\xi_m) \right\} \leq h(q),$$

which contradicts (4.46) and hence the claim. From (4.49)

$$\varphi(x_0, t_0) - \varphi(\xi_{s_1}(s_1), s_1) = \int_{s_1}^{t_0} \frac{\mathrm{d}}{\mathrm{d}\theta} \varphi(\dot{\xi}_{s_1}(\theta), \theta)\mathrm{d}\theta$$

$$= \int_{s_1}^{t_0} (\varphi_t + \langle \dot{\xi}_{s_1}, D\varphi\rangle)(\xi_{s_1}(\theta), \theta) d\theta$$
$$\leq \int_{s_1}^{t_0} g(\xi_{s_1}) - \epsilon(t_0 - s_1),$$

which contradicts (4.47). This proves that $\underline{u}$ is a super solution and hence it is a viscosity solution. Similarly from Lemma 4.8, it follows that $\bar{u}$ is a viscosity solution. This together with Lemma 4.6 completes the proof of the Theorem.

Remark 4.9. In Lemma 4.6, assumptions on g are only sufficient but not necessary. For example consider the problem

$$u_t + \mathrm{e}^{-u}|u_x| = g(t)$$
$$u(x,0) = u_0(x).$$

Solutions of this problem are given by

$$\underline{u}(x,t) = \inf_y \left\{ u_0(y) + \int_0^t g(s)\mathrm{d}s;\ u_0(y) \leq -\log\left(\frac{|x-y|}{\int_0^t \exp(-\int_0^s g(\theta)\mathrm{d}\theta)\mathrm{d}s}\right)\right\} \tag{4.52}$$

and

$$\bar{u}(x,t) = \inf_y \left\{ u_0(y) + \int_0^t g(s)\mathrm{d}s;\ u_0(y) < -\log\left(\frac{|x-y|}{\int_0^t \exp(-\int_0^s g(\theta)\mathrm{d}\theta)\mathrm{d}s}\right)\right\}. \tag{4.53}$$

Furthermore $\underline{u}^* = \bar{u}$ and $\bar{u}_* = \underline{u}$. Here $g(t) \leq 0$ is not required.

Proof. By formula

$$\underline{u}(x,t) = \inf_{\xi \in C(x,t)} \left\{ u_0(y) + \int_0^t g(s)\mathrm{d}s;\ u_0(y) \leq \operatorname*{ess\,inf}_{0\leq s\leq t} \left\{ h(\dot{\xi}(s)) - \int_0^s g(\theta)\mathrm{d}\theta \right\}\right\},$$

where $h(q) = \log\left(\frac{1}{|q|}\right)$. Let

$$\bar{\xi}(s) = \left(\frac{x-y}{\int_0^t \exp(-\int_0^s g(\theta)\mathrm{d}\theta)\mathrm{d}s}\right)\left(\int_0^s \exp\left(-\int_0^\theta g(\eta)\mathrm{d}\eta\right)\mathrm{d}\theta\right) + y.$$

Then $\bar{\xi}(t) = x, \bar{\xi}(0) = y$ and $\dot{\xi}(s) = ((x-y)/(\int_0^t \exp(-\int_0^s g(\theta)\mathrm{d}\theta)\mathrm{d}s)) \exp(-\int_0^s g(\eta)\mathrm{d}\eta)$. Also

$$h(\dot{\xi}(s)) - \int_0^s g(\eta)\mathrm{d}\eta = -\log\left(\frac{|x-y|}{\int_0^t \exp(-\int_0^s g(\theta)\mathrm{d}\theta)\mathrm{d}s}\right).$$

Therefore

$$\underline{u}(x,t) \leq \inf_y \left\{ u_0(y) + \int_0^t g(s)\mathrm{d}s;\ u_0(y) \leq -\log\left(\frac{|x-y|}{\int_0^t \exp(-\int_0^s g(\theta)\mathrm{d}\theta)\mathrm{d}s}\right)\right\}.$$

On the other hand,

$$u_0(y) \le h(\dot{\xi}(\theta)) - \int_0^{\theta} g(\eta)\mathrm{d}\eta \quad \forall\, \theta \in [0,t]$$

implies

$$\exp\left(-u_0(y) - \int_0^{\theta} g(\eta)\mathrm{d}\eta\right) \ge |\dot{\xi}(\theta)|.$$

On integration over $[0,t]$, we have since $\int_0^t |\dot{\xi}(\theta)| \ge |x-y|$,

$$u_0(y) \le -\log\left(\frac{|x-y|}{\int_0^t \exp(-\int_0^s g(\theta)\mathrm{d}\theta)\mathrm{d}s}\right).$$

This implies

$$\underline{u}(x,t) \ge \inf_y \left\{ u_0(y) + \int_0^t g(s)\mathrm{d}s;\ u_0(y) \le -\log\left(\frac{|x-y|}{\int_0^t \exp(-\int_0^s g(\theta)\mathrm{d}\theta)\mathrm{d}s}\right)\right\}.$$

Hence (4.52). Similarly (4.53) follows.

Choose $M_0 > 0$ be such that $\int_{t_1}^t g(\eta)\mathrm{d}\eta \le M_0(t-t_1)$ for all $t_1 < t$. Since $\int_0^t \exp(-\int_0^s g(\theta)\,\mathrm{d}\theta)\mathrm{d}s$ is an increasing function of t, it follows that

$$\begin{aligned}
\underline{u}(x,t_1) &= \inf_y \left\{ u_0(y) + \int_0^{t_1} g(s)\mathrm{d}s;\ u_0(y) \le -\log\left(\frac{|x-y|}{\int_o^{t_1} \exp(-\int_0^s g(\theta)\mathrm{d}\theta)\mathrm{d}s}\right)\right\} \\
&\ge \inf_y \left\{ u_0(y) + \int_0^t g(s)\mathrm{d}s - \int_{t_1}^t g(s)\mathrm{d}s;\right. \\
&\qquad\qquad \left. u_0(y) < -\log\left(\frac{|x-y|}{\int_0^t \exp(-\int_0^s g(\theta)\mathrm{d}\theta)\mathrm{d}s}\right)\right\} \\
&\ge \overline{u}(x,t) - M_0(t-t_1).
\end{aligned}$$

Therefore

$$\begin{aligned}
\underline{u}^*(x,t) &\ge \lim_{t_1 \to t} \underline{u}(x,t_1) \\
&\ge \lim_{t_1 \to t} (\overline{u}(x,t) - M_0(t-t_1)) \\
&= \overline{u}(x,t).
\end{aligned}$$

Hence we have $\underline{u}^* = \overline{u}$ and similarly $\overline{u}_* = \underline{u}$.

Acknowledgement

The authors would like to thank the referee for his critical comments and for pointing out a mistake and improvement of Theorem 2.2 of an earlier version of this paper. The authors acknowledge funding from Indo-French Centre for Promotion of Advanced Research under Project 1901–2.

References

[1] Adimurthi and Veerappa Gowda G D, Hopf Lax type formula for sub and supersolutions, *Adv. Diff. Eq.* **5** (2000) 97–119

[2] Barron E N, Evans L C and Jensen R, Viscosity solutions of Isaacs equations and differential games with Lipschitz controls, *J. Diff. Eq.* **53** (1984) 213–233

[3] Barron E N and Ishii H, The Bellman equation for minimizing maximum cost, *Nonlinear Analysis TMA* **13(9)** (1989) 1067–1090

[4] Barron E N and Liu W, Calculus of variations in L^{∞}, *Appl. Math. Opt.* **35** (1997) 237–263

[5] Barron E N, Jensen R and Liu W, Hopf-Lax type formula for $u_t + H(u, Du) = 0$, *J. Diff. Eq.* **126** (1996) 48–64

[6] Barles G and Perthame B, Discontinuous solutions of deterministic optimal stopping time problems, *Math. Modelling Numer. Anal.* **21** (1987) 57–579

[7] Evans L C, *Partial differential equations*, Berkeley Mathematics Lecture Notes, (1994) Vols 3A, 3B

[8] Evans L C and Souganidis P, Differential games and representation formulas for Hamilton Jacobi equations, *Indiana Univ. Math. J.* **33** (1984) 773–795

[9] Hitoshi Ishii, Perron's method for Hamilton Jacobi equations, *Duke. Math. J.* **55** (1987) 369–364

[10] Lions P L, *Generalized solutions of Hamilton Jacobi equations*, Research notes in Mathematics (Pitmann) (1982)

[11] Rudin W, *Functional Analysis* (Tata McGraw Hill Pub.) (1974)

Proc. Indian Acad. Sci. (Math. Sci.), Vol. 110, No. 4, November 2000, pp. 415–430.
© Printed in India

Completely monotone multisequences, symmetric probabilities and a normal limit theorem

J C GUPTA

Indian Statistical Institute, 7, SJS Sansanwal Marg, New Delhi 110016, India
Corresponding Address: 32, Mirdha Tola, Budaun 243 601, India

MS received 20 October 1999; revised 7 April 2000

Abstract. Let $\mathbf{G}_{n,k}$ be the set of all partial completely monotone multisequences of order n and degree k, i.e., multisequences $c_n(\beta_1,\beta_2,\ldots,\beta_k)$, $\beta_1,\beta_2,\ldots,\beta_k=0,1,2,\ldots$, $\beta_1+\beta_2+\cdots+\beta_k\le n, c_n(0,0,\ldots,0)=1$ and $(-1)^{\beta_0}\Delta^{\beta_0}c_n(\beta_1,\beta_2,\ldots,\beta_k)\ge 0$ whenever $\beta_0\le n-(\beta_1+\beta_2+\cdots+\beta_k)$ where $\Delta c_n(\beta_1,\beta_2,\ldots,\beta_k)=c_n(\beta_1+1,\beta_2,\ldots,\beta_k)+c_n(\beta_1,\beta_2+1,\ldots,\beta_k)+\cdots+c_n(\beta_1,\beta_2,\ldots,\beta_k+1)-c_n(\beta_1,\beta_2,\ldots,\beta_k)$. Further, let $\prod_{n,k}$ be the set of all symmetric probabilities on $\{0,1,2,\ldots,k\}^n$. We establish a one-to-one correspondence between the sets $\mathbf{G}_{n,k}$ and $\prod_{n,k}$ and use it to formulate and answer interesting questions about both. Assigning to $\mathbf{G}_{n,k}$ the uniform probability measure, we show that, as $n\to\infty$, any fixed section $\{c_n(\beta_1,\beta_2,\ldots,\beta_k)$, $1\le\sum\beta_i\le m\}$, properly centered and normalized, is asymptotically multivariate normal. That is, $\left\{\sqrt{\binom{n+k}{k}}(c_n(\beta_1,\beta_2,\ldots,\beta_k)-c_0(\beta_1,\beta_2,\ldots,\beta_k),\ 1\le\beta_1+\beta_2+\cdots+\beta_k\le m\right\}$ converges weakly to MVN$[0,\Sigma_m]$; the centering constants $c_0(\beta_1,\beta_2,\ldots,\beta_k)$ and the asymptotic covariances depend on the moments of the Dirichlet $(1,1,\ldots,1;1)$ distribution on the standard simplex in R^k.

Keywords. Moment spaces; completely monotone multisequences; normal limit.

1. Introduction

The notion of a *completely monotone multisequence* was introduced in [2] in the context of solving the moment problem on the standard k-simplex. We recall the definition.

DEFINITION

A multisequence of real numbers

$$c(\beta_1,\beta_2,\ldots,\beta_k),\beta_1,\beta_2,\ldots,\beta_k=0,1,2,\ldots,c(0,0,\ldots,0)=1 \tag{1.1}$$

is said to be *completely monotone* if

$$(-1)^{\beta_0}\Delta^{\beta_0}c(\beta_1,\beta_2,\ldots,\beta_k)\ge 0\ \forall\ \beta_1,\beta_2,\ldots,\beta_k=0,1,2,\ldots, \tag{1.2}$$

where

$$\begin{aligned}\Delta c(\beta_1,\beta_2,\ldots,\beta_k)=&\,c(\beta_1+1,\beta_2,\ldots,\beta_k)+c(\beta_1,\beta_2+1,\ldots,\beta_k)+\cdots\\&+c(\beta_1,\beta_2,\ldots,\beta_k+1)-c(\beta_1,\beta_2,\ldots,\beta_k)\end{aligned} \tag{1.3}$$

and Δ^{β_0} stands for β_0 iterates of the operator Δ.

The following theorem solves the moment problem on the standard k-simplex

$$S_k := \{(x_1,x_2,\dots,x_k) : x_i \geq 0, x_1+x_2+\cdots+x_k \leq 1\}. \tag{1.4}$$

Theorem 1.1. [2]. *There exists a probability measure P on S_k such that*

$$\int_{S_k} x_1^{\beta_1}x_2^{\beta_2}\cdots x_k^{\beta_k}\mathrm{d}P = c(\beta_1,\beta_2,\dots,\beta_k) \tag{1.5}$$

for all $\beta_1,\beta_2,\dots,\beta_k = 0,1,2,\dots$ if and only if the multisequence $\{c(\beta_1,\beta_2,\dots,\beta_k) : \beta_1,\beta_2,\dots,\beta_k = 0,1,2,\dots\}$ is completely monotone.

For $k=1, S_k=[0,1]$, a completely monotone multisequence is just a completely monotone sequence and the above theorem reduces to Hausdorff's solution [4] to the moment problem on [0, 1]. In the sequel, for the sake of brevity, we call a completely monotone sequence an H-sequence and a competely monotone multisequence a G-multisequence.

In [3] we explored the interplay between partial H-sequences and symmetric probabilities on products of $\{0,1\}$. In response to a query raised by a referee of [2] and [3], in this paper we carry out an analysis of the interplay between G-multisequences on the one hand and symmetric probabilities on products of $\{0,1,2,\dots,k\}$ on the other. This paper is organized as follows. In § 2 we establish an affine bijection between the set of all partial G-multisequences of order n and degree k (see below for definition) and the set of all symmetric probabilities on $\{0,1,2,\dots,k\}^n$ and use it to answer questions about both these sets. In § 3 we formulate and prove a normal limit theorem for partial G-multisequences. In § 4 we make some remarks.

2. G-multisequences and symmetric probabilities

We introduce the notion of a partial G-multisequence of order n and degree k.

DEFINITION

A multisequence

$$\begin{aligned} &c_n(\beta_1,\beta_2,\dots,\beta_k),\ \beta_1,\beta_2,\dots,\beta_k = 0,1,2,\dots \\ &c_n(0,0,\dots,0)=1,\ \beta_1+\beta_2+\cdots+\beta_k \leq n \end{aligned} \tag{2.1}$$

such that

$$(-1)^{\beta_0}\Delta^{\beta_0}c_n(\beta_1,\beta_2,\dots,\beta_k) \geq 0 \ \forall \ \beta_0 \leq n-\beta_1-\beta_2-\cdots-\beta_k \tag{2.2}$$

is called a partial G-multisequence of order n and degree k. The set

$$\begin{aligned} \mathbf{G}_{n,k} :=&\{\{c_n(\beta_1,\beta_2,\dots,\beta_k), 1 \leq \beta_1+\beta_2+\cdots+\beta_k \leq n\} : \\ &(-1)^{\beta_0}\Delta^{\beta_0}c_n(\beta_1,\beta_2,\dots,\beta_k) \geq 0 \\ &\text{for } \beta_0 = 0,1,\dots,n-\beta_1-\beta_2-\cdots-\beta_k\} \end{aligned} \tag{2.3}$$

with the understanding that $c_n(0,0,\dots,0) \equiv 1$ denotes the set of all partial G-multisequences of order n and degree k.

Notation. k is a natural number; $\alpha_i,\beta_i,\gamma_i,\delta_i, i=1,2,\dots,k$, are non-negative integers and $x_1,x_2,\dots,x_k$ are real numbers.

For $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_k)$,

$$\begin{aligned}
|\boldsymbol{\beta}| &:= \beta_1 + \beta_2 + \cdots + \beta_k;\\
\mathbf{x}^{\boldsymbol{\beta}} &:= \prod_{i=1}^{k} x_i^{\beta_i} \text{ where } \mathbf{x} = (x_1, x_2, \ldots, x_k);\\
L_{m,k} &:= \{(\beta_1, \beta_2, \ldots, \beta_k) : |\boldsymbol{\beta}| = m\}, \quad m = 0, 1, 2, \ldots;\\
\mathcal{L}_{j,k} &:= \bigcup_{m=0}^{j} L_{m,k} = \{\boldsymbol{\beta} : |\boldsymbol{\beta}| \le j\};\\
\tilde{\mathcal{L}}_{j,k} &:= \bigcup_{m=1}^{j} L_{m,k} = \{\boldsymbol{\beta} : 1 \le |\boldsymbol{\beta}| \le j\};\\
\boldsymbol{\beta}! &:= \prod_{i=1}^{k} \beta_i!;\\
\binom{n}{\boldsymbol{\beta}} &:= \frac{n!}{\prod_{i=1}^{k} \beta_i! (n - |\boldsymbol{\beta}|)!}.
\end{aligned} \tag{2.4}$$

We say $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_k) \le \boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_k)$ if $\beta_i \le \gamma_i, i = 1, 2, \ldots, k$ and in that case define

$$\binom{\boldsymbol{\gamma}}{\boldsymbol{\beta}} := \frac{\boldsymbol{\gamma}!}{\boldsymbol{\beta}!(\boldsymbol{\gamma} - \boldsymbol{\beta})!}.$$

We define $\boldsymbol{\alpha} \vee \boldsymbol{\beta} := (\alpha_1 \vee \beta_1, \alpha_2 \vee \beta_2, \ldots, \alpha_k \vee \beta_k)$. For $\boldsymbol{\delta} = \boldsymbol{\alpha} + \boldsymbol{\beta} + \boldsymbol{\gamma}$, we define

$$\binom{\boldsymbol{\delta}}{\boldsymbol{\alpha}\ \boldsymbol{\beta}\ \boldsymbol{\gamma}} := \prod_{i=1}^{k} \binom{\delta_i}{\alpha_i\ \beta_i\ \gamma_i}.$$

For $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_k)$ and $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_k)$,

$$\sum_{\boldsymbol{\lambda}=0}^{\boldsymbol{\gamma}} \text{ stands for } \sum_{\lambda_1=0}^{\gamma_1} \sum_{\lambda_2=0}^{\gamma_2} \cdots \sum_{\lambda_k=0}^{\gamma_k}.$$

For a partial multisequence $c_n(\boldsymbol{\beta})$ of order n, we define, for $|\boldsymbol{\beta}| \le m \le n$,

$$q_m(\boldsymbol{\beta}) = (-1)^{m-|\beta|} \Delta^{m-|\beta|} c_n(\boldsymbol{\beta}) \tag{2.5}$$

and observe that, by (2.2),

$$q_m(\boldsymbol{\beta}) \ge 0, \quad |\boldsymbol{\beta}| \le m \le n. \tag{2.6}$$

For $m \ge |\boldsymbol{\beta}| + 1$, we define

$$\nabla q_m(\boldsymbol{\beta}) = q_m(\boldsymbol{\beta}) + q_m(\beta_1 + 1, \beta_2, \ldots, \beta_k) + \cdots + q_m(\beta_1, \beta_2, \ldots, \beta_k + 1). \tag{2.7}$$

By (2.5) and (2.7), it easily follows that

$$\nabla q_m(\boldsymbol{\beta}) = q_{m-1}(\boldsymbol{\beta}), \quad |\boldsymbol{\beta}| + 1 \le m \le n. \tag{2.8}$$

Consequently, for $|\boldsymbol{\beta}| \leq m \leq n$,

$$q_m(\boldsymbol{\beta}) = \nabla^{n-m} q_n(\boldsymbol{\beta}) = \sum_{j=0}^{n-m} \binom{n-m}{j} \sum_{\delta \in L_{j,k}} \binom{j}{\boldsymbol{\delta}} q_n(\boldsymbol{\beta}+\boldsymbol{\delta})$$

$$= \sum_{\delta \in \mathcal{L}_{n-m,k}} \binom{n-m}{\boldsymbol{\delta}} q_n(\boldsymbol{\beta}+\boldsymbol{\delta}). \tag{2.9}$$

By (2.5), for $\boldsymbol{\beta} \in \mathcal{L}_{n,k}$,

$$c_n(\boldsymbol{\beta}) = q_{|\beta|}(\boldsymbol{\beta}) = \sum_{\delta \in \mathcal{L}_{n-|\beta|,k}} \binom{n-|\boldsymbol{\beta}|}{\boldsymbol{\delta}} q_n(\boldsymbol{\beta}+\boldsymbol{\delta}). \tag{2.10}$$

By (2.5), for $\boldsymbol{\beta} \in \mathcal{L}_{n,k}$,

$$q_n(\boldsymbol{\beta}) = (-1)^{n-|\beta|} \Delta^{n-|\beta|} c_n(\boldsymbol{\beta}) = \sum_{\delta \in \mathcal{L}_{n-|\beta|,k}} (-1)^{|\delta|} \binom{n-|\boldsymbol{\beta}|}{\boldsymbol{\delta}} c_n(\boldsymbol{\beta}+\boldsymbol{\delta}). \tag{2.11}$$

By (2.10),

$$\sum_{\beta \in \mathcal{L}_{n,k}} \binom{n}{\boldsymbol{\beta}} q_n(\boldsymbol{\beta}) = \nabla^n q_n(0) = c_n(0) = 1. \tag{2.12}$$

Finally, by (2.9), the non-negativity of $q_n(\boldsymbol{\beta})$'s implies that conditions (2.2) hold and consequently,

$$\mathbf{G}_{n,k} = \{\{c_n(\boldsymbol{\beta}), \boldsymbol{\beta} \in \tilde{\mathcal{L}}_{n,k}\} : q_n(\boldsymbol{\beta}) \geq 0 \ \forall \ \boldsymbol{\beta} \in \tilde{\mathcal{L}}_{n,k}\}, \tag{2.13}$$

where $q_n(\boldsymbol{\beta})$'s are given by (2.11). Of course $c_n(0) = 1$.

Let $E_k := \{0, 1, 2, \ldots, k\}$. Given an element $\{c_n(\boldsymbol{\beta}), \boldsymbol{\beta} \in \tilde{\mathcal{L}}_{n,k}\}$ of $\mathbf{G}_{n,k}$, let Q_n be the symmetric measure on the n-fold product space E_k^n which assigns mass $q_n(\boldsymbol{\beta})$ to each point $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$ of E_k^n for which $\#\{i : \omega_i = j\} = \beta_j$, $j = 1, 2, \ldots, k$ and $\#\{i : \omega_i = 0\} = n - |\boldsymbol{\beta}|$. By (2.6) and (2.12), Q_n is a probability. Conversely, given $\{q_n(\boldsymbol{\beta}), \boldsymbol{\beta} \in \tilde{\mathcal{L}}_{n,k}\}$, eqs (2.10) determine an element of $\mathbf{G}_{n,k}$. This establishes a one-to-one correspondence between $\mathbf{G}_{n,k}$ and

$$\prod_{n,k} := \left\{ \{q_n(\boldsymbol{\beta}), \boldsymbol{\beta} \in \mathcal{L}_{n,k}\} : q_n(\boldsymbol{\beta}) \geq 0, \sum_{\beta \in \tilde{\mathcal{L}}_{n,k}} \binom{n}{\boldsymbol{\beta}} q_n(\boldsymbol{\beta}) \leq 1 \right\}, \tag{2.14}$$

the set of all symmetric probabilities on E_k^n. Of course

$$q_n(0) = 1 - \sum_{\beta \in \tilde{\mathcal{L}}_{n,k}} \binom{n}{\boldsymbol{\beta}} q_n(\boldsymbol{\beta}). \tag{2.15}$$

Equations (2.10) and (2.11) define maps

$$\phi_n : \mathbf{G}_{n,k} \longrightarrow \prod_{n,k}$$

and

$$\psi_n : \prod_{n,k} \longrightarrow \mathbf{G}_{n,k}. \tag{2.16}$$

Clearly, these maps establish an affine congruence between the convex sets $\mathbf{G}_{n,k}$ and $\prod_{n,k}$. Further, these maps are inverses of each other.

We define the projection map

$$\pi_n : \mathbf{G}_{n+1,k} \longrightarrow \mathbf{G}_{n,k}$$

by

$$\{c(\boldsymbol{\beta}) : \boldsymbol{\beta} \in \tilde{\mathcal{L}}_{n+1,k}\} \longmapsto \{c(\boldsymbol{\beta}) : \boldsymbol{\beta} \in \tilde{\mathcal{L}}_{n,k}\}. \tag{2.17}$$

Likewise, we define

$$\tilde{\pi}_n : \prod_{n+1,k} \longmapsto \prod_{n,k}$$

by

$$q^* \longmapsto q, \tag{2.18}$$

where q is the n-dimensional marginal of q^* in $\prod_{n+1,k}$. We note that both these projection maps are affine. We briefly illustrate the use of these maps to answer questions about $\mathbf{G}_{n,k}$ and $\prod_{n,k}$.

(a) *Extreme points of* $\mathbf{G}_{n,k}$ *and* $\prod_{n,k}$. Clearly, the extreme points of $\prod_{n,k}$ correspond to probabilities $Q_n^{\boldsymbol{\gamma}}, \boldsymbol{\gamma} \in \mathcal{L}_{n,k}$, where $Q_n^{\boldsymbol{\gamma}}$ is the uniform probability distribution on the set of those elements of E_k^n for which $\#\{i : w_i = j\} = \gamma_j, j = 1, 2, \ldots, k$, i.e.,

$$q_n^{\boldsymbol{\gamma}}(\boldsymbol{\beta}) = \begin{cases} \dfrac{1}{\binom{n}{\gamma}} & \text{if} \quad \boldsymbol{\beta} = \boldsymbol{\gamma} \\ 0 & \text{otherwise.} \end{cases} \tag{2.19}$$

The extreme points of $\mathbf{G}_{n,k}$, which are otherwise not so apparent, can easily be obtained by using the map ψ_n which, being an affine congruence, maps $\partial\prod_{n,k}$ onto $\partial\mathbf{G}_{n,k}$. Simple calculations show that

$$\partial\mathbf{G}_{n,k} = \{\{c_n^{\boldsymbol{\gamma}}(\boldsymbol{\beta}), \boldsymbol{\beta} \in \tilde{\mathcal{L}}_{n,k}\} : \boldsymbol{\gamma} \in \mathcal{L}_{n,k}\},$$

where

$$c_n^{\boldsymbol{\gamma}}(\boldsymbol{\beta}) = \begin{cases} \dfrac{\binom{\gamma}{\beta}}{\binom{n}{\beta}} = \dfrac{\prod_{i=1}^{k}(\gamma_i(\gamma_i - 1) \cdots (\gamma_i - \beta_i + 1))}{n(n-1) \cdots (n - |\beta| + 1)} & \text{if } \boldsymbol{\beta} \leq \boldsymbol{\gamma} \\ 0 & \text{otherwise.} \end{cases} \tag{2.20}$$

(b) *Extendability of partial* $\mathbf{G}$*-multisequences.* We define

$$\mathbf{G}_{n,k}^{n+m,k} := \left\{ \begin{matrix} \{c(\boldsymbol{\beta}), \boldsymbol{\beta} \in \tilde{\mathcal{L}}_{m,k}\} : \exists\{c(\boldsymbol{\beta}), \boldsymbol{\beta} \in \tilde{\mathcal{L}}_{n+m,k} - \tilde{\mathcal{L}}_{n,k}\} \text{ s.t.} \\ \{c(\boldsymbol{\beta}) : \boldsymbol{\beta} \in \tilde{\mathcal{L}}_{n+m,k}\} \in \mathbf{G}_{n+m,k} \end{matrix} \right\}. \tag{2.21}$$

This is the set of those partial multisequences of order n which can be extended to ones of order $n + m$. Clearly,

$$\mathbf{G}_{n,k}^{n+m,k} = \pi_n \circ \pi_{n+1} \circ \cdots \circ \pi_{n+m-1}(\mathbf{G}_{n+m,k})$$

and consequently,

$$\mathbf{G}_{n,k}^{n+m,k} = \text{Convex Hull } \{\pi_n \circ \pi_{n+1} \circ \ldots \circ \pi_{n+m-1}(\partial \mathbf{G}_{n+m,k})\}.$$

Simple calculations show that

$$\partial \mathbf{G}_{n,k}^{n+m,k} = \{\{c_n^{\gamma}(\boldsymbol{\beta}), \boldsymbol{\beta} \in \tilde{\mathcal{L}}_{n,k}\} : \boldsymbol{\gamma} \in \mathcal{L}_{n+m,k}\},$$

where

$$c_n^{\gamma}(\boldsymbol{\beta}) = \begin{cases} \dfrac{\binom{\gamma}{\beta}}{\binom{n+m}{\beta}} = \dfrac{\prod_{i=1}^{k}(\gamma_i(\gamma_i-1)\cdots(\gamma_i-\beta_i+1))}{(n+m)(n+m-1)\cdots(n+m-|\beta|+1)} & \text{if } \boldsymbol{\beta} \leq \boldsymbol{\gamma} \\ 0 & \text{otherwise.} \end{cases} \tag{2.22}$$

(c) *Extendability of symmetric probabilities.* We define

$$\prod_{n,k}^{n+m,k} := \left\{ \begin{matrix} \{q(\boldsymbol{\beta}), \boldsymbol{\beta} \in \tilde{\mathcal{L}}_{n,k}\} : \exists \{q(\boldsymbol{\beta}), \boldsymbol{\beta} \in \tilde{\mathcal{L}}_{n+m,k} - \tilde{\mathcal{L}}_{n,k}\} \\ \text{s.t. } \{q(\boldsymbol{\beta}), \boldsymbol{\beta} \in \tilde{\mathcal{L}}_{n+m,k}\} \in \prod_{n+m,k} \end{matrix} \right\}. \tag{2.23}$$

This is the set of those symmetric probabilities on E_k^n which can be extended to symmetric probabilities on E_k^{n+m}. Simple calculations show that

$$\partial \prod_{n,k}^{n+m,k} = \{\{q_n^{\gamma}(\boldsymbol{\beta}), \boldsymbol{\beta} \in \tilde{\mathcal{L}}_{n,k}\} : \boldsymbol{\gamma} \in \mathcal{L}_{n+m,k}\},$$

where

$$q_n^{\gamma}(\boldsymbol{\beta}) = \begin{cases} \dfrac{\binom{m}{\gamma-\beta}}{\binom{n+m}{\gamma}} & \text{if } \boldsymbol{\beta} \leq \boldsymbol{\gamma} \text{ and } |\boldsymbol{\beta}| \geq |\boldsymbol{\gamma}| - m \\ 0 & \text{otherwise.} \end{cases} \tag{2.24}$$

(d) *The moment space of the standard k-simplex.* We define

$$\mathbf{M_{n,k}} = \bigcap_{m=1}^{\infty} \mathbf{G}_{n,k}^{n+m,k} = \bigcap_{m=1}^{\infty} [\text{Convex Hull } \{\{c_n^{\gamma}(\boldsymbol{\beta}), \boldsymbol{\beta} \in \tilde{\mathcal{L}}_{n,k}\} : \boldsymbol{\gamma} \in \mathcal{L}_{n+m,k}\}], \tag{2.25}$$

where $c_n^{\gamma}(\cdot)$ are given by (2.22). This is the set of those partial multisequences of order n which can be extended to ones of order $m+n$ for each $m = 1, 2, \ldots$. In view of Theorem 1.1, it follows that

$$\mathbf{M_{n,k}} = \left\{ \left\{ c_n(\boldsymbol{\beta}) = \int_{S_k} x^{\beta} d\lambda, \ \boldsymbol{\beta} \in \tilde{\mathcal{L}}_{n,k} \right\} : \lambda \in \Lambda \right\}, \tag{2.26}$$

where Λ is the set of all probability measures on the simplex S_k.

Equation (2.25) expresses the nth moment space, i.e., the space of all moments of total order n or less, of the simplex, as an intersection of a decreasing sequence of polytopes in $R^{|\tilde{\mathcal{L}}_{n,k}|}$. Each $\mathbf{G}_{n,k}^{n+m,k}$ can be written as an intersection of finitely many closed half-spaces and consequently, $\mathbf{M_{n,k}}$ can be expressed as an intersection of countably many closed half-spaces of $R^{|\tilde{\mathcal{L}}_{n,k}|}$.

The above description of $\mathbf{M_{n,k}}$ is rather indirect and we do not consider it very useful. Kemperman and Skibinsky [6], among others, gave explicit description of $\mathbf{M_{n,k}}$ for $n = 2, 3$ and $k = 2$; their description involves some non-linear conditions.

3. A normal limit theorem for G-multisequences

To get a better insight into the shape and structure of the sets $\mathbf{G}_{n,k}$, we would like to look at a typical point of it. For this purpose, for each $n = 1, 2, \ldots$, we put uniform probability measure on $\mathbf{G}_{n,k}$, i.e., the Lebesgue measure normalized by volume $(\mathbf{G}_{n,k})$. The coordinate variables $c_n(\boldsymbol{\beta})$ can now be viewed as random variables on the probability space $\mathbf{G}_{n,k}$. We fix an initial segment $\{c_n(\boldsymbol{\beta}), \boldsymbol{\beta} \in \tilde{\mathcal{L}}_{m,k}\}$, and prove that, after suitable centering and normalization, it converges in law, as $n \to \infty$, to a multivariate normal (briefly: MVN). This convergence to MVN is *controlled* by the Dirichlet distribution $D = D(1, 1, \ldots 1; 1)$ on the simplex S_k in the sense that the centering constants and the covariance matrix of the limiting MVN depend on the moments of D; see (3.12) and (3.13).

We find it convenient to introduce a suitable system of canonical coordinates in the space $\mathbf{G}_{n,k}$. Let

$$\mathcal{S}_{n,k} = \left\{ \{p_n(\boldsymbol{\beta}), \boldsymbol{\beta} \in \tilde{\mathcal{L}}_{n,k}\} : p_n(\boldsymbol{\beta}) \geq 0, \sum_{\beta \in \tilde{\mathcal{L}}_{n,k}} p_n(\boldsymbol{\beta}) \leq 1 \right\} \tag{3.1}$$

be the standard simplex in R^N, where

$$N = N(n, k) = |\tilde{\mathcal{L}}_{n,k}| = \binom{n+k}{k} - 1. \tag{3.2}$$

We put

$$p_n(0) = 1 - \sum_{\beta \in \tilde{\mathcal{L}}_{n,k}} p_n(\boldsymbol{\beta}) \tag{3.3}$$

and set up a bijection between $\mathcal{S}_{n,k}$ and $\prod_{n,k}$ by putting

$$p_n(\boldsymbol{\beta}) = \binom{n}{\boldsymbol{\beta}} q_n(\boldsymbol{\beta}), \quad \boldsymbol{\beta} \in \tilde{\mathcal{L}}_{n,k}. \tag{3.4}$$

Of course, by (2.15) and (3.4),

$$p_n(0) = q_n(0). \tag{3.5}$$

This gives a bijection between $\mathbf{G}_{n,k}$ and $\mathcal{S}_{n,k}$. Explicitly, by (2.10), (2.11) and (3.4), we have

$$c_n(\boldsymbol{\beta}) = \frac{1}{\binom{n}{\beta}} \sum_{\delta \in \mathcal{L}_{n-|\beta|,k}} \binom{\boldsymbol{\beta}+\boldsymbol{\delta}}{\boldsymbol{\beta}} p_n(\boldsymbol{\beta}+\boldsymbol{\delta}) \tag{3.6}$$

and

$$p_n(\boldsymbol{\beta}) = \binom{n}{\boldsymbol{\beta}} \sum_{\delta \in \mathcal{L}_{n-|\beta|,k}} (-1)^{|\delta|} \binom{n-|\boldsymbol{\beta}|}{\boldsymbol{\delta}} c_n(\boldsymbol{\beta}+\boldsymbol{\delta}). \tag{3.7}$$

We shall employ $\{p_n(\boldsymbol{\beta}), \boldsymbol{\beta} \in \tilde{\mathcal{L}}_{n,k}\}$ as the canonical coordinates of the space $\mathbf{G}_{n,k} \subset R^{N(n,k)}$. We take $\{e(\boldsymbol{\beta}), \boldsymbol{\beta} \in \tilde{\mathcal{L}}_{n,k}\}$ as the standard basis in $R^{N(n,k)}$ and order the elements of this basis by assigning the natural order to $L_{j,k}, j = 1, 2, \ldots, n$ and within each $L_{j,k}$ choosing and fixing some linear order. Clearly, the matrices of linear transformations (3.6) and (3.7) with respect to the chosen ordered basis, are upper triangular. Further,

$$\frac{\partial(c_n(\boldsymbol{\beta}), \boldsymbol{\beta} \in \tilde{\mathcal{L}}_{n,k})}{\partial(p_n(\boldsymbol{\beta}), \boldsymbol{\beta} \in \tilde{\mathcal{L}}_{n,k})} = \left[\prod_{\beta \in \tilde{\mathcal{L}}_{n,k}} \binom{n}{\boldsymbol{\beta}}\right]^{-1}. \qquad (3.8$$

We need some combinatorial identities.

PROPOSITION 3.1

(a) $$\sum_{j=0}^{t} \binom{s+j}{s} = \binom{s+t+1}{s+1} \textit{ for } s, t = 0, 1, 2, \ldots. \qquad (3.9a$$

(b) $$\sum_{j=0}^{t} \binom{-a}{t-j}\binom{-b}{j} = \binom{-(a+b)}{t} \textit{ for } t = 0, 1, 2, \ldots \textit{ and}$$

real a and b, or equivalently,

$$\sum_{j=0}^{t} \binom{a+t-j-1}{t-j}\binom{b+j-1}{j} = \binom{a+b+t-1}{t}. \qquad (3.9b$$

(c) $$\sum_{\gamma \in L_{j,k}} \binom{\beta+\gamma}{\beta} = \binom{|\beta|+k+j-1}{j} \qquad (3.9c$$

for $j = 0, 1, 2, \ldots, k = 1, 2, \ldots$ *and* $\beta = (\beta_1, \beta_2, \ldots, \beta_k), \beta_i = 0, 1, 2, \ldots.$

(d) $$\sum_{\gamma \in \mathcal{L}_{t,k}} \binom{\beta+\gamma}{\beta} = \binom{|\beta|+k+t}{t} \textit{ for } k = 1, 2, \ldots, t = 0, 1, 2, \ldots. \qquad (3.9d$$

(e) $$\sum_{\delta \in L_{t,k}} \prod_{i=1}^{k} (1+\rho_i)^{\delta_i} = \sum_{r=0}^{t} \binom{t+k-1}{r+k-1} \sum_{\gamma \in \mathcal{L}_{r,k}} \prod_{i=1}^{k} \rho_i^{\gamma_i} \qquad (3.9e$$

for $t = 0, 1, 2, \ldots$ *and* $k = 1, 2, \ldots.$

(f) $$\prod_{\beta \in \mathcal{L}_{n,k}} \beta! = \prod_{r=0}^{n} (r!)^{k\binom{n-r+k-1}{k-1}} \qquad (3.9$$

for $n = 0, 1, 2, \ldots, k = 1, 2, \ldots.$

(g) $$\sum_{\delta \in L_{t,k}} \binom{\gamma+\delta}{\alpha}\binom{\gamma+\delta}{\beta}$$

$$= \sum_{r=o}^{t\wedge(|\alpha|+|\beta|)} \sum_{\lambda=0}^{\gamma} \binom{\gamma}{\lambda} \sum_{\substack{\phi \in L_{r,k} \\ \gamma \leq \phi+\lambda \leq \alpha+\beta}} \binom{t+k-1}{r+k-1} \binom{\phi+\lambda}{\phi+\lambda-\beta \ \phi+\lambda-\alpha \ \alpha+\beta-\phi-\lambda}$$

(3.9

for $\alpha, \beta \in \mathcal{L}_{m,k}$ *and* $\gamma = \alpha \vee \beta.$

Proof. (a) This is easily proved by induction. □

(b) This follows by comparing the coefficients of x^j in the identity $(1+x)^{-a}(1+x)^{-b} = (1+x)^{-(a+b)}$. □

(c) The case $k=1$ is trivial. Fix $k \geq 2$; we use induction. Assume that the identity holds for all j, all $\boldsymbol{\beta}$ and k replaced by $k-1$. Then

$$\begin{aligned}\sum_{\gamma \in L_{j,k}} \binom{\boldsymbol{\beta}+\boldsymbol{\gamma}}{\boldsymbol{\beta}} &= \sum_{i=0}^{j} \binom{\beta_1+i}{\beta_1} \sum_{\gamma_2+\gamma_3+\cdots+\gamma_k=j-i} \prod_{\lambda=2}^{k} \binom{\beta_\lambda+\gamma_\lambda}{\beta_\lambda} \\ &= \sum_{i=0}^{j} \binom{\beta_1+i}{\beta_1} \binom{|\boldsymbol{\beta}|-\beta_1+k-1+j-i-1}{j-i} \\ &\qquad \text{by induction hypothesis} \\ &= \binom{|\boldsymbol{\beta}|+k+j-1}{j} \text{ by (b).}\end{aligned}$$ □

(d) This follows from (c) and (a). □

(e) The identity trivially holds for $t=0, k=1,2\ldots$ and for $k=1, t=1,2,\ldots$. We give the lexicographic ordering to the index set $I=\{(s,\lambda) : s=0,1,2,\ldots,\lambda=1,2,\ldots\}$. Fix $(t,k), t\geq 1$ and $k \geq 2$ and assume the induction hypothesis that the equality holds for all elements of I strictly preceding (t,k). We prove the equality for (t,k). The term $\rho_1^{\gamma_1}\prod_{i=2}^{k}\rho_i^{\gamma_i}$, with $\gamma \in L_{r,k}$, occurs in

$$\sum_{\delta\in L_{t,k}} \prod_{i=1}^{k} (1+\rho_i)^{\delta_i} = \sum_{\delta \in L_{t,k}} (1+\rho_1)^{\delta_1} \prod_{i=2}^{k} (1+\rho_i)^{\delta_i}$$

with coefficient

$$\begin{aligned}\sum_{\delta_1=\gamma_1}^{t-r+\gamma_1} \binom{\delta_1}{\gamma_1}\binom{t-\delta_1+k-2}{r-\gamma_1+k-2} &= \sum_{m=0}^{t-r} \binom{\gamma_1+m}{\gamma_1}\binom{t-\gamma_1-m+k-2}{r-\gamma_1+k-2} \\ &= \sum_{m=0}^{t-r}\binom{\gamma_1+1+(m-1)}{m} \\ &\quad \cdot \binom{r-\gamma_1+k-1+(t-r-m-1)}{t-r-m} \\ &= \binom{t+k-1}{t-r} \text{ by (3.9b).} \\ &= \binom{t+k-1}{r+k-1}.\end{aligned}$$

Thus

$$\sum_{\delta\in L_{t,k}} \prod_{i=1}^{k}(1+\rho_i)^{\delta_i} = \sum_{r=0}^{t} \binom{t+k-1}{r+k-1} \sum_{\gamma\in L_{r,k}} \prod_{i=1}^{k} \rho_i^{\gamma_i}.$$ □

(f) The identity trivially holds for $n=0, k=1,2,\ldots$ and $k=1, n=1,2,\ldots$. We give the lexicographic ordering to the index set $I=\{(m,\lambda) : m=0,1,2,\ldots,\lambda=1,2,\ldots\}$.

Fix $(n,k), n \geq 1$ and $k \geq 2$ and assume the induction hypothesis that the equality holds for all elements of I strictly preceding (n,k). We prove the equality for (n,k). We have

$$\prod_{\beta \in \mathcal{L}_{n,k}} \boldsymbol{\beta}! = \prod_{r=0}^{n} \prod_{\beta_2+\beta_3+\cdots+\beta_k \leq n-r} \{r!(\beta_2!\beta_3!\ldots\beta_k!\}$$

$$= \left\{\prod_{r=0}^{n}(r!)^{|\mathcal{L}(n-r,k-1)|}\right\}\left\{\prod_{r=0}^{n}\prod_{\beta \in \mathcal{L}_{n-r,k-1}} \boldsymbol{\beta}!\right\}$$

$$= \left\{\prod_{r=0}^{n}(r!)^{|\mathcal{L}(n-r,k-1)|}\right\}\left\{\prod_{r=0}^{n}\prod_{s=0}^{n-r}(s!)^{(k-1)\binom{n-r-s+k-2}{k-2}}\right\}$$

by the induction hypothesis

$$= \left\{\prod_{r=0}^{n}(r!)^{|\mathcal{L}(n-r,k-1)|}\right\}\left\{\prod_{s=0}^{n}(s!)^{(k-1)\sum_{r=0}^{n-s}\binom{n-r-s+k-2}{k-2}}\right\}$$

$$= \prod_{r=0}^{n}(r!)^{k\binom{n-r+k-1}{k-1}} \text{ by (3.9a).}$$

□

(g) Let $\mathbf{x} = (x_1, x_2, \ldots, x_k), \mathbf{y} = (y_1, y_2, \ldots, y_k)$ and $\mathbf{1} + \boldsymbol{\rho} := (1+\rho_1, 1+\rho_2, \ldots, 1+\rho_k$ with $\rho_i = x_i + y_i + x_i y_i$. The LHS of the identity can be identified with the coefficien of $\mathbf{x}^\alpha \mathbf{y}^\beta$ in

$$\sum_{\delta \in L_{t,k}} [(\mathbf{1}+\mathbf{x})(\mathbf{1}+\mathbf{y})]^{\gamma+\delta} = (\mathbf{1}+\boldsymbol{\rho})^\gamma \sum_{\delta \in L_{t,k}} (\mathbf{1}+\boldsymbol{\rho})^\delta$$

$$= (\mathbf{1}+\boldsymbol{\rho})^\gamma \sum_{r=0}^{t} \binom{t+k-1}{r+k-1} \sum_{\phi \in L_{r,k}} \boldsymbol{\rho}^\phi \text{ by (3.9e)}$$

$$= \sum_{r=0}^{t}\sum_{\lambda=0}^{\gamma} \binom{\boldsymbol{\gamma}}{\boldsymbol{\lambda}} \sum_{\phi \in L_{r,k}} \binom{t+k-1}{r+k-1} \boldsymbol{\rho}^{\phi+\lambda}.$$

We now observe that the coefficient of $u^a v^b$ in $(u+v+uv)^m$ equals $\left(\begin{matrix} & m & \\ m-b & m-a & a+b-m \end{matrix}\right.$ if $a \vee b \leq m \leq a+b$ and zero otherwise. Thus

$$\text{LHS} = \sum_{r=0}^{t}\sum_{\lambda=0}^{\gamma} \binom{\boldsymbol{\gamma}}{\boldsymbol{\lambda}} \sum_{\substack{\phi \in L_{r,k} \\ \gamma \leq \phi+\lambda \leq \alpha+\beta}} \binom{t+k-1}{r+k-1} \left(\begin{matrix} & \boldsymbol{\phi}+\boldsymbol{\lambda} & \\ \boldsymbol{\phi}+\boldsymbol{\lambda}-\boldsymbol{\beta} & \boldsymbol{\phi}+\boldsymbol{\lambda}-\boldsymbol{\alpha} & \boldsymbol{\alpha}+\boldsymbol{\beta}-\boldsymbol{\phi}-\boldsymbol{\lambda} \end{matrix}\right)$$

We now observe that if $r > |\boldsymbol{\alpha}| + |\boldsymbol{\beta}|$ then $|\boldsymbol{\phi}| + |\boldsymbol{\lambda}| > |\boldsymbol{\alpha}| + |\boldsymbol{\beta}| + |\boldsymbol{\lambda}|$ and $\gamma_i \leq \phi_i + \lambda_i$ $\alpha_i + \beta_i$ is violated for some i. Thus LHS of the identity equals

$$\sum_{r=0}^{t \wedge (|\alpha|+|\beta|)} \sum_{\lambda=0}^{\gamma} \binom{\boldsymbol{\gamma}}{\boldsymbol{\lambda}} \sum_{\substack{\phi \in L_{r,k} \\ \gamma \leq \phi+\lambda \leq \alpha+\beta}} \binom{t+k-1}{r+k-1} \left(\begin{matrix} & \boldsymbol{\phi}+\boldsymbol{\lambda} & \\ \boldsymbol{\phi}+\boldsymbol{\lambda}-\boldsymbol{\beta} & \boldsymbol{\phi}+\boldsymbol{\lambda}-\boldsymbol{\alpha} & \boldsymbol{\alpha}+\boldsymbol{\beta}-\boldsymbol{\phi}-\boldsymbol{\lambda} \end{matrix}\right)$$

□

We put

$$V_{n,k} = \text{Volume}\,(\mathbf{G}_{n,k}). \tag{3.10}$$

PROPOSITION 3.2

As $n \to \infty$,

$$\frac{1}{n^{k+1}} \log V_{n,k} \to -A_k$$

where

$$A_k = \frac{k(k+1)}{k!} \sum_{j=0}^{k-1} \frac{(-1)^j \binom{k-1}{j}}{(j+2)^2}. \tag{3.11}$$

The constants A_k are positive with $A_1 = \frac{1}{2}$, $A_2 = \frac{5}{12}$, $A_3 = \frac{13}{72}, \ldots$; for $k = 1$ the above result agrees with the corresponding one in [3].

Proof.

$$\begin{aligned} V_{n,k} &= \int_{\mathbf{G}_{n,k}} \prod_{\beta \in \tilde{\mathcal{L}}_{n,k}} \mathrm{d}c_n(\boldsymbol{\beta}) \\ &= \left[\prod_{\beta \in \tilde{\mathcal{L}}_{n,k}} \binom{n}{\boldsymbol{\beta}} \right]^{-1} \int_{\mathcal{S}_{n,k}} \prod_{\beta \in \tilde{\mathcal{L}}_{n,k}} \mathrm{d}p_n(\boldsymbol{\beta}) \text{ by (3.8)} \\ &= \frac{\{\prod_{\beta \in \tilde{\mathcal{L}}_{n,k}} \boldsymbol{\beta}!\}\{\prod_{j=1}^{n} [(n-j)!]^{|L_{j,k}|}\}}{N(n,k)!(n!)^{|\tilde{\mathcal{L}}_{n,k}|}} \text{ by (2.4)} \\ &= \frac{\prod_{r=1}^{n} (r!)^{(k+1)\binom{n-r+k-1}{k-1}}}{N(n,k)!(n!)^{\binom{n+k}{k}}} \text{ by (3.2) and (3.9f).} \end{aligned}$$

Now,

$$\log N(n,k)! = o(n^{k+1}),$$

$$\binom{n+k}{k} \log n! = \frac{n^{k+1}}{k!} \log n - \frac{n^{k+1}}{k!} + o(n^{k+1})$$

and

$$\begin{aligned} \log\left\{ \prod_{r=1}^{n} (r!)^{(k+1)\binom{n-r+k-1}{k-1}} \right\} &= (k+1)\, \frac{n^{k-1}}{k-1!} \sum_{r=1}^{n} \left(1 - \frac{r}{n}\right)^{k-1} (r \log r - r) + o(n^{k+1}) \\ &= \frac{(k+1)n^{k+1}}{k-1!} \sum_{r=1}^{n} \frac{1}{n} \left(1 - \frac{r}{n}\right)^{k-1} \left[\frac{r}{n} \log \frac{r}{n} + \frac{r}{n} \log n - \frac{r}{n}\right] + o(n^{k+1}) \\ &= \frac{(k+1)n^{k+1}}{k-1!} \left[\int_0^1 x(1-x)^{k-1} \log x \,\mathrm{d}x + \log n \int_0^1 x(1-x)^{k-1} \mathrm{d}x \right. \\ &\quad \left. - \int_0^1 x(1-x)^{k-1} \mathrm{d}x \right] + o(n^{k+1}). \end{aligned}$$

Thus

$$\log V_{n,k} = \frac{k(k+1)}{k!} \cdot n^{k+1} \int_0^1 x(1-x)^{k-1} \log x \, \mathrm{d}x + o(n^{k+1})$$

$$= -\frac{k(k+1)}{k!} \cdot n^{k+1} \cdot \sum_{j=0}^{k-1} \frac{(-1)^j \binom{k-1}{j}}{(j+2)^2} + o(n^{k+1}). \qquad \square$$

PROPOSITION 3.3

The uniform probability measure on the space $\mathbf{G}_{n,k}$ *is equivalent to having the uniform probability measure on the space* $\mathcal{S}_{n,k}$ *of canonical coordinates.*

Proof. This is an immediate consequence of (3.8) and the change of variables formula for an integral on $\mathbf{G}_{n,k}$ to $\mathcal{S}_{n,k}$ and vice-versa. □

The main result of this section is the following.

Theorem 3.4. *For each* $m = 1, 2, \ldots,$ *as* $n \to \infty$, *the law of*

$$\left\{ \sqrt{\binom{n+k}{k}} (c_n(\boldsymbol{\beta}) - c_0(\boldsymbol{\beta})),\ \boldsymbol{\beta} \in \tilde{\mathcal{L}}_{m,k} \right\}$$

relative to uniform probability on $\mathbf{G}_{n,k}$ *converges weakly to a multivariate normal distribution MVN* $[0, \Sigma_m]$ *on* $R^{N(m,k)}$, *where*

$$c_0(\boldsymbol{\beta}) = \frac{k!\boldsymbol{\beta}!}{(k+|\boldsymbol{\beta}|)!}, \quad \Sigma_m = ((\sigma(\boldsymbol{\alpha}, \boldsymbol{\beta})))_{\alpha,\beta \in \tilde{\mathcal{L}}_{m,k}}$$

$$\sigma(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{(\boldsymbol{\alpha}+\boldsymbol{\beta})!k!}{(|\boldsymbol{\alpha}| + |\boldsymbol{\beta}| + k)!}. \tag{3.12}$$

We note that the centering constants and the covariances are given by the moments of the Dirichlet distribution $D(1, 1, \ldots; 1)$ on S_k:

$$c_0(\boldsymbol{\beta}) = \mu(\boldsymbol{\beta}), \quad \sigma(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \mu(\boldsymbol{\alpha} + \boldsymbol{\beta}),$$

where

$$\mu(\boldsymbol{\gamma}) = \int_{S_k} x_1^{\gamma_1} x_2^{\gamma_2} \cdots x_k^{\gamma_k} \, \mathrm{d}x_1 \, \mathrm{d}x_2 \cdots \mathrm{d}x_k. \tag{3.13}$$

Proof. The uniform probability on the simplex $\mathcal{S}_{n,k}$ is just the Dirichlet $(1, 1, \ldots, 1; 1)$ distribution on it. Let $Z_0, Z_1 \ldots$ be a sequence of i.i.d. standard exponential random variables defined on, say, the probability space $(\Omega, \mathcal{F}, P)$. We relabel these random variables as Z_α, $\boldsymbol{\alpha} \in \bigcup_{n=0}^{\infty} L_{n,k}$. The law, under P, of $\left\{ \frac{Z(\beta)}{\sum_{\alpha \in \mathcal{L}_{n,k}} Z(\alpha)} : \boldsymbol{\beta} \in \tilde{\mathcal{L}}_{n,k} \right\}$ is Dirichlet $(1, 1, \ldots; 1)$. Hence, by (3.6) and Proposition 3.3, the law of $\{c_n(\boldsymbol{\beta}), \boldsymbol{\beta} \in \tilde{\mathcal{L}}_{n,k}\}$, under uniform probability on $\mathbf{G}_{n,k}$, is same as the law, under P, of

$$Y_n(\boldsymbol{\beta})=\left[\sum_{\delta\in\mathcal{L}_{n-|\beta|}}\binom{\boldsymbol{\beta}+\boldsymbol{\delta}}{\boldsymbol{\delta}}Z(\boldsymbol{\beta}+\boldsymbol{\delta})\right]\Bigg/\left[\binom{n}{\boldsymbol{\beta}}\sum_{\alpha\in\mathcal{L}_{n,k}}Z(\boldsymbol{\alpha})\right],\quad \boldsymbol{\beta}\in\tilde{\mathcal{L}}_{n,k}. \tag{3.14}$$

Now choose and fix an integer $m\geq 1$ and consider $n\geq m$. We observe that, for $\boldsymbol{\beta}\in\tilde{\mathcal{L}}_{m,k}$,

$$E(Y_n(\boldsymbol{\beta}))=\left[\sum_{\delta\in\mathcal{L}_{n-|\beta|}}\binom{\boldsymbol{\beta}+\boldsymbol{\delta}}{\boldsymbol{\delta}}\right]\Bigg/\left[\binom{n}{\boldsymbol{\beta}}\binom{n+k}{k}\right]=\binom{k+|\boldsymbol{\beta}|}{k}^{-1}\text{ by (3.9d)}$$

$$=c_0(\boldsymbol{\beta})\text{ by (3.12).}$$

We further observe that

$$\left[\sum_{\alpha\in\mathcal{L}_{n,k}}Z(\boldsymbol{\alpha})\right]\Bigg/\left[\binom{n+k}{k}\right]\overset{P}{\rightarrow}1. \tag{3.15}$$

Hence, by (3.14) and (3.15), to prove the stated weak convergence of $\left\{\sqrt{\binom{n+k}{k}}(c_n(\boldsymbol{\beta})-c_0(\boldsymbol{\beta})),\boldsymbol{\beta}\in\tilde{\mathcal{L}}_{m,k}\right\}$, it suffices to prove that

$$\{T_n(\boldsymbol{\beta}),\boldsymbol{\beta}\in\tilde{\mathcal{L}}_{m,k}\}\Longrightarrow\text{MVN}(0,\Sigma_m),$$

where

$$T_n(\boldsymbol{\beta}):=\left[\sum_{\delta\in\mathcal{L}_{n-|\beta|,k}}\binom{\boldsymbol{\beta}+\boldsymbol{\delta}}{\boldsymbol{\delta}}\{Z(\boldsymbol{\beta}+\boldsymbol{\delta})-1\}\right]\Bigg/\left[\binom{n}{\boldsymbol{\beta}}\sqrt{\binom{n+k}{k}}\right]. \tag{3.16}$$

For $\boldsymbol{\alpha},\boldsymbol{\beta}\in\tilde{\mathcal{L}}_{m,k}$, Let $\boldsymbol{\gamma}=\boldsymbol{\alpha}\vee\boldsymbol{\beta}$. Then, for $n\geq|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|+|\boldsymbol{\gamma}|$, we have

$$\text{Cov}\,(T_n(\boldsymbol{\alpha}),T_n(\boldsymbol{\beta}))=\left[\sum_{\delta\in\mathcal{L}_{n-|\gamma|,k}}\binom{\boldsymbol{\gamma}+\boldsymbol{\delta}}{\boldsymbol{\alpha}}\binom{\boldsymbol{\gamma}+\boldsymbol{\delta}}{\boldsymbol{\beta}}\right]\Bigg/\left[\binom{n}{\boldsymbol{\alpha}}\binom{n}{\boldsymbol{\beta}}\binom{n+k}{k}\right]$$

and, by (3.9g),

$$\left[\binom{n}{\boldsymbol{\alpha}}\binom{n}{\boldsymbol{\beta}}\binom{n+k}{k}\right]\text{Cov}\,(T_n(\boldsymbol{\alpha}),T_n(\boldsymbol{\beta}))=\sum_{t=0}^{n-|\gamma|}\sum_{r=0}^{t\wedge|\alpha+\beta|}\sum_{\lambda=0}^{\gamma}\binom{\boldsymbol{\gamma}}{\boldsymbol{\lambda}}$$

$$\times\sum_{\substack{\phi\in L_{r,k}\\ \gamma\leq\phi+\lambda\leq\alpha+\beta}}\binom{t+k-1}{r+k-1}\binom{\boldsymbol{\phi}+\boldsymbol{\lambda}}{\boldsymbol{\phi}+\boldsymbol{\lambda}-\boldsymbol{\beta}\ \ \boldsymbol{\phi}+\boldsymbol{\lambda}-\boldsymbol{\alpha}\ \ \boldsymbol{\alpha}+\boldsymbol{\beta}-\boldsymbol{\phi}-\boldsymbol{\lambda}}$$

$$=\sum_{r=0}^{|\alpha+\beta|}\sum_{\lambda=0}^{\gamma}\binom{\boldsymbol{\gamma}}{\boldsymbol{\lambda}}\sum_{\substack{\phi\in L_{r,k}\\ \gamma\leq\phi+\lambda\leq\alpha+\beta}}\binom{\boldsymbol{\phi}+\boldsymbol{\lambda}}{\boldsymbol{\phi}+\boldsymbol{\lambda}-\boldsymbol{\beta}\ \ \boldsymbol{\phi}+\boldsymbol{\lambda}-\boldsymbol{\alpha}\ \ \boldsymbol{\alpha}+\boldsymbol{\beta}-\boldsymbol{\phi}-\boldsymbol{\lambda}}\sum_{t=r}^{n-|\gamma|}\binom{t+k-1}{r+k-1}$$

$$=\sum_{r=0}^{|\alpha+\beta|}\sum_{\lambda=0}^{\gamma}\binom{\boldsymbol{\gamma}}{\boldsymbol{\lambda}}\sum_{\substack{\phi\in L_{r,k}\\ \gamma\leq\phi+\lambda\leq\alpha+\beta}}\binom{\boldsymbol{\phi}+\boldsymbol{\lambda}}{\boldsymbol{\phi}+\boldsymbol{\lambda}-\boldsymbol{\beta}\ \ \boldsymbol{\phi}+\boldsymbol{\lambda}-\boldsymbol{\alpha}\ \ \boldsymbol{\alpha}+\boldsymbol{\beta}-\boldsymbol{\phi}-\boldsymbol{\lambda}}\binom{n-|\boldsymbol{\gamma}|+k}{r+k}$$

by (3.9a).

As $n \to \infty$, the dominating term corresponds to $r = |\boldsymbol{\alpha} + \boldsymbol{\beta}|$ and the conditions $r = |\boldsymbol{\alpha} + \boldsymbol{\beta}|, \boldsymbol{\phi} \in L_{r,k}, \boldsymbol{\gamma} \le \boldsymbol{\phi} + \boldsymbol{\lambda} \le \boldsymbol{\alpha} + \boldsymbol{\beta}$ imply that $\boldsymbol{\phi} = \boldsymbol{\alpha} + \boldsymbol{\beta}$ and $\boldsymbol{\lambda} = 0$. Thus

$$\begin{aligned} \operatorname{Cov}(T_n(\boldsymbol{\alpha}), T_n(\boldsymbol{\beta})) &\sim \left[\binom{n}{\boldsymbol{\alpha}}\binom{n}{\boldsymbol{\beta}}\binom{n+k}{k}\right]^{-1} \binom{n - |\boldsymbol{\gamma}| + k}{|\boldsymbol{\alpha}| + |\boldsymbol{\beta}| + k}\binom{\boldsymbol{\alpha}+\boldsymbol{\beta}}{\boldsymbol{\alpha}} \\ &\sim \frac{(\boldsymbol{\alpha}+\boldsymbol{\beta})!k!}{(|\boldsymbol{\alpha}| + |\boldsymbol{\beta}| + k)!} = \sigma(\boldsymbol{\alpha}, \boldsymbol{\beta}) \text{ by (3.12).} \end{aligned} \tag{3.17}$$

By the Cramér–Wold theorem it suffices to prove that

$$\sum_{\beta \in \tilde{\mathcal{L}}_{m,k}} a(\boldsymbol{\beta}) T_n(\boldsymbol{\beta}) \Longrightarrow N\left[0, \sum\sum a(\boldsymbol{\alpha})a(\boldsymbol{\beta})\sigma(\boldsymbol{\alpha}, \boldsymbol{\beta})\right]$$

for all $\{a(\boldsymbol{\beta}), \boldsymbol{\beta} \in \tilde{\mathcal{L}}_{m,k}\}$. We have

$$\begin{aligned} &\sum_{\beta \in \tilde{\mathcal{L}}_{m,k}} a(\boldsymbol{\beta}) T_n(\boldsymbol{\beta}) \\ &= \binom{n+k}{k}^{-1/2} \left[\sum_{\beta \in \tilde{\mathcal{L}}_{m,k}} a(\boldsymbol{\beta}) \binom{n}{\boldsymbol{\beta}}^{-1} \sum_{\delta \in \mathcal{L}_{n-|\beta|,k}} \binom{\boldsymbol{\beta}+\boldsymbol{\delta}}{\boldsymbol{\delta}} (Z(\boldsymbol{\beta}+\boldsymbol{\delta}) - 1)\right] \\ &= \binom{n+k}{k}^{-1/2} \sum_{\gamma \in \mathcal{L}_{n,k}} b_n(\boldsymbol{\gamma})(Z(\boldsymbol{\gamma}) - 1), \end{aligned} \tag{3.18}$$

where

$$b_n(\boldsymbol{\gamma}) = \sum_{\substack{\beta \le \gamma \\ \beta \in \tilde{\mathcal{L}}_{m,k}}} \frac{a(\boldsymbol{\beta})\binom{\gamma}{\beta}}{\binom{n}{\beta}}. \tag{3.19}$$

We write

$$U_n := \sum_{\gamma \in \mathcal{L}_{n,k}} X_n(\boldsymbol{\gamma}) \text{ with } X_n(\boldsymbol{\gamma}) = b_n(\boldsymbol{\gamma})(Z(\boldsymbol{\gamma}) - 1)$$

and observe that U_n is a sum of independent random variables centered at their expectations. Further, we have the following:

(i)
$$\begin{aligned} s_n^2 = \operatorname{Var}(U_n) &= \binom{n+k}{k} \operatorname{Var}\left(\sum a(\boldsymbol{\beta}) T_n(\boldsymbol{\beta})\right) \\ &\sim \binom{n+k}{k} \sum\sum a(\boldsymbol{\alpha})a(\boldsymbol{\beta})\sigma(\boldsymbol{\alpha}, \boldsymbol{\beta}) \text{ by (3.17)} \end{aligned}$$

(ii)
$$\begin{aligned} \frac{1}{s_n^3} \sum_{\gamma \in \mathcal{L}_{n,k}} E|X_n(\boldsymbol{\gamma})|^3 &= \frac{1}{s_n^3} \sum_{\gamma \in \mathcal{L}_{n,k}} |b_n(\boldsymbol{\gamma})|^3 E|Z(\boldsymbol{\gamma}) - 1|^3 \\ &\to 0, \quad \text{as } n \to \infty, \end{aligned}$$

since $|b_n(\boldsymbol{\gamma})| \le \sum_{\beta \in \tilde{\mathcal{L}}_{m,k}} |a(\boldsymbol{\beta})|$ by (3.19), $|\mathcal{L}_{n,k}| = \binom{n+k}{k}$ and $s_n^3 = O\left(\binom{n+k}{k}^{3/2}\right)$ by (i).

Hence, by Loéve [7], p. 275, $U_n/s_n \Longrightarrow N(0,1)$, or equivalently,

$$\sum_{\beta\in\tilde{\mathcal{L}}_{m,k}} a(\boldsymbol{\beta})T_n(\boldsymbol{\beta}) = \binom{n+k}{k}^{-1/2} U_n \Longrightarrow N\left[0, \sum\sum a(\boldsymbol{\alpha})a(\boldsymbol{\beta})\sigma(\boldsymbol{\alpha},\boldsymbol{\beta})\right].$$

This completes the proof. □

4. Remarks

The moment spaces $\mathbf{M_n}$ of $[0,1]$ – in our notation $\mathbf{M_n} = \mathbf{M_{n,1}}$ – have been comprehensively treated in the literature. Karlin and McGregor [5] study a random walk $\{X_n\}_{n\geq 0}$ on the nonnegative integers Z_+ such that

$$P(X_{n+1} = j|X_n = i) = \begin{cases} p_i & \text{if } j = i-1 \\ 1-p_i & \text{if } j = i+1 \\ 0 & \text{otherwise} \end{cases} \tag{4.1}$$

and $0 < p_i < 1$ for $i \geq 1$, while $p_0 = 0$; they show that there exists a unique probability measure λ of infinite support on $[0,1]$ such that

$$P_{00}^{(2n)} = P(X_{2n} = 0|X_0 = 0) = \int_0^1 x^n \,\mathrm{d}\lambda \text{ for all } n \geq 0. \tag{4.2}$$

This essentially establishes a one-to-one correspondence between the moment sequences on one hand and the parameters $\{p_i\}_{i\geq 0}$ of the associated random walks on the other; taking $0 < p_i < 1$ amounts to confining oneself to the interior of $\mathbf{M_n}$. The parameters $p_i, i \geq 1$, associated with λ turn out to be canonical moments of λ introduced by Skibinsky [8]. Chang, Kemperman and Studden [1] employ $\{p_i\}$ as the canonical coordinates of the space $\mathbf{M_n}$ and prove a normal limit theorem for moment sequences.

All our efforts to find analogues of the random walk of Karlin and McGregor [5] or of canonical moments associated with a probability measure on S_k were unsuccessful. It would be desirable to do an asymptotic analysis of the moment spaces $\mathbf{M_{n,k}}$ for $k \geq 2$. That, in the present state of our knowledge, appears to be a formidable task.

Acknowledgements

The author thanks Morris Skibinsky for bringing ref. [6] to his attention and the Indian Statistical Institute, New Delhi Center, for the facilities extended to him. The author also thanks an anonymous referee for a careful reading of the manuscript and making suggestions for improvement in the presentation.

References

[1] Chang F C, Kemperman J H B and Studden W J, A normal limit theorem for moment sequences. *Ann. Probab.* **21** (1993) 1295–1309

[2] Gupta J C, The moment problem for the standard *k*-dimensional simplex. *Sankhyā, Series A* **61** (1999) 286–291

[3] Gupta J C, Partial Hausdorff sequences and symmetric probabilities on finite products of $\{0,1\}$. *Sankhyā, Series A* **61** (1999) 347–357

[4] Hausdorff F, Momentprobleme für ein endliches Intervall, *Math. Z.* **16** (1923) 220–248

[5] Karlin S and McGregor J, Random walks, *Illinois J. Math.* **3** (1959) 66–81

[6] Kemperman J H B and Skibinsky M, Covariance spaces for measures on polyhedral sets. In *Stochastic Inequalities* (eds) M Shaked, and Y L Tong, *IMS Lecture Notes* **22** (1993) 182–195

[7] Loéve M, *Probability Theory*, 3rd edition, (New York: Van Nostrand) (1963)

[8] Skibinsky M, The range of $(r+1)$th moment for distributions on $[0,1]$, *J. Appl. Probab.* **4** (1967) 543–552

Proc. Indian Acad. Sci. (Math. Sci.), Vol. 110, No. 4, November 2000, pp. 431–447.
© Printed in India

An asymptotic derivation of weakly nonlinear ray theory

PHOOLAN PRASAD

Department of Mathematics, Indian Institute of Science, Bangalore 560 012, India
Email: prasad@math.iisc.ernet.in

MS received 7 January 2000; revised 28 April 2000

Abstract. Using a method of expansion similar to Chapman–Enskog expansion, a new formal perturbation scheme based on high frequency approximation has been constructed. The scheme leads to an eikonal equation in which the leading order amplitude appears. The transport equation for the amplitude has been deduced with an error $O(\epsilon^2)$ where ϵ is the small parameter appearing in the high frequency approximation. On a length scale over which Choquet–Bruhat's theory is valid, this theory reduces to the former. The theory is valid on a much larger length scale and the leading order terms give the weakly nonlinear ray theory (WNLRT) of Prasad, which has been very successful in giving physically realistic results and also in showing that the caustic of a linear theory is resolved when nonlinear effects are included. The weak shock ray theory with infinite system of compatibility conditions also follows from this theory.

Keywords. Nonlinear wave propagation; ray theory; hyperbolic equations; caustic.

1. Introduction

A ray theory is a result of a mathematical method of finding an approximate value of the solution of a hyperbolic system of partial differential equations based on the high frequency approximation. The high frequency approximation implies that we can formally distinguish between the amplitude w and a phase function $\phi(\mathbf{x}, t)$ whose level surfaces in $\mathbf{x}$-space (at a fixed t) define a one parameter family of wavefronts. In a ray theory we study the successive positions of a given wavefront and also attempt to calculate the amplitude distribution w on it. The method, when applied to a linear system of equations, gives rise to linear rays whose equations decouple from the transport equation for the amplitude. Quite frequently, the linear rays starting from the points of a curved wavefront envelop a caustic surface on which the linear wavefront has cusp type of singularities where the assumptions of the ray theory break down. Thus the ray theory applied to a linear system is valid only over a distance which is small compared to the distance of an arête (where caustic begins to form) from the initial position of the wavefront. For a linear wavefront propagating in an uniform isotropic medium at rest, this distance R of arête is equal to the minimum of the principal radii of curvature.

Experimental results [20] and theoretical investigations [14, 18] have shown that the amplitude of a wavefront, even when small amplitude assumption is made, has significant effect on rays and the wavefront geometry. Hence a nonlinear ray theory requires derivation of two equations, the first one being the eikonal equation

$$Q(\nabla\phi, \phi_t, \mathbf{x}, t, w(\mathbf{x}, t)) = 0, \quad \nabla = (\partial_{x_1}, \ldots, \partial_{x_n}) \tag{1.1}$$

for the phase function ϕ such that the amplitude w of the wavefront appears in the eikonal equation itself. The second equation is a transport equation for the amplitude along a nonlinear ray. The nonlinear rays are curves $\mathbf{x} = \mathbf{x}(t)$ obtained from the solution of the characteristic equations or Hamilton's canonical equations of (1.1). Though, ray theories for the propagation of a nonlinear wavefront were dealt by many [7, 22, 10, 11]; Choquet-Bruhat [4] presented a systematic formal derivation of it for a general hyperbolic system of quasilinear partial differential equations. Hunter and his collaborators extended Choquet-Bruhat's theory to many important situations [5]. The Choquet-Bruhat's perturbation procedure leads to an eikonal equation which is independent of the amplitude w and therefore, uses a transport equation for the amplitude along linear rays. Naturally the theory is valid over a distance over which a linear theory is valid i.e. on a length scale much smaller than R. Over a number of years, Prasad [12–15] (see also [17, 18]) has developed a nonlinear ray theory in which the eikonal equation depends also on the amplitude and the transport equation is along the nonlinear rays. The solution influences the wavefront geometry in such a way that the radii of curvature has no relevance as a length scale in the problem. Both experimental [20] and theoretical results [14, 18] show that for a moderately weak nonlinear wave, the caustic does not appear in the solution. The wavefront geometry consists of almost plane segments joined by kinks across which the amplitude and the ray direction suffer jump. On each of these segments the amplitude of the wave varies slowly. Therefore, in a moderately weak nonlinear ray theory, the problem is not to find the solution in a caustic region because the caustic itself does not appear but to find the new geometry of a nonlinear wavefront and the solution on it. It turns out that except for immediate neighbourhoods of the kinks, the nonlinear ray theory is valid on a much larger length scale. The structure of the kinks can be studied on a smaller length scale by two-dimensional Burger's equations (or Zaboltskaya–Khokhlov or Z–K equation) which have been studied by Hunter and his coworkers [2, 3, 6] and Tabak and Rosales [21].

The aim of this paper is to construct a formal perturbation scheme which leads to an eikonal equation in which the leading order amplitude w appears and to derive a transport equation for w along the corresponding nonlinear rays. We have been able to deduce the transport equation for w with an error of the order of ϵ^2. The method of expansion is similar to the Chapman–Enskog expansion, a discussion of which for a hyperbolic system is available in an article by Hunter [5]. A careful examination of the various terms in ray equations and the transport equation show that in practice only a few terms may be retained and this leads to the nonlinear ray theory of Prasad, which has been very successful in giving physically realistic results and also in showing that the caustic of a linear theory is resolved when nonlinear effects are included [18, 9]. These two references contain extensive numerical results of the approximate equations derived in this paper. So does the paper of Kevlahan who shows that the shock ray theory derived in the end of the § 5 gives results which agrees well with experimental results, known expressions for approximate solutions and numerical solution of full Euler equations. A still better comparison with numerical solution of Euler equations is being worked out but this will take some time and will be reported later.

2. An asymptotic derivation of WNLRT

We consider a hyperbolic system of first order quasilinear partial differential equations

$$A(\mathbf{u})\mathbf{u}_t + B^{(\alpha)}(\mathbf{u})\mathbf{u}_{\mathbf{x}_\alpha} = 0 \qquad \alpha = 1, 2, \ldots, m. \tag{2.1}$$

Here, $\mathbf{x} \in R^m$ are the space variables, $\mathbf{u}(\mathbf{x}, t) \in R^n$ are the dependent variables , and $A(\mathbf{u})$ and $B^{(\alpha)}(\mathbf{u})$, are smooth $n \times n$ matrix-valued functions of $\mathbf{u}$. We use the summation convention over repeated indices. We only consider smooth solutions, so it is not necessary to write the system in conservation form.

We look for a generalized asymptotic expansion of solutions of (2.1) of the following form:

$$\mathbf{u}(\mathbf{x}, t, \epsilon) = \epsilon\mathbf{v}\left(\mathbf{x}, t, \frac{\phi(\mathbf{x}, t, \epsilon)}{\epsilon}, \epsilon\right), \tag{2.2}$$

$$\mathbf{v}(\mathbf{x}, t, \theta, \epsilon) = \mathbf{v}_0(\mathbf{x}, t, \theta, \epsilon) + \epsilon\mathbf{v}'(\mathbf{x}, t, \theta, \epsilon), \quad \theta = \phi/\epsilon. \tag{2.3}$$

Here ϵ is a small parameter, so this ansatz represents a small amplitude high frequency solution. The function $\phi(\mathbf{x}, t, \epsilon) \in R$ and the functions $\mathbf{v}_0(\mathbf{x}, t, \theta, \epsilon)$ and $\mathbf{v}'(\mathbf{x}, t, \theta, \epsilon)$ will be chosen so that (2.2) gives an asymptotic solution of (2.1) as $\epsilon \to 0$. In particular, ϕ is the phase function associated with the leading order solution $\mathbf{u} = \epsilon\mathbf{v}_0$. When carrying out the expansion, we assume that the derivatives, ϕ_{x_α}, are of order one with respect to ϵ.

This method of expansion is similar to the Chapman–Enskog expansion. For other work in nonlinear hyperbolic waves using Chapman–Enskog expansions see [5]. The leading order solution $\mathbf{v}_0$ and the correction $\mathbf{v}'$ depend on ϵ explicitly. It is therefore not necessary to include any higher order terms in the expansion (2.3), since they can be absorbed into $\mathbf{v}'$. As a result of this explicit ϵ dependence, the solution $\mathbf{v}$ can be decomposed into a leading order approximation, $\mathbf{v}_0$ and a perturbation $\epsilon\mathbf{v}'$ in different ways, since terms in $\mathbf{v}'$ can be absorbed into $\mathbf{v}_0$. One way to specify the decomposition uniquely is to require that

$$\mathbf{l}_0 \cdot \mathbf{v}' = 0, \tag{2.4}$$

where the left null vector $\mathbf{l}_0$ is defined below. However, other choices are possible. For example in gas dynamics we could require that $\mathbf{v}'$ contains no pressure perturbations.

We now derive the asymptotic equations. We will obtain an asymptotic solution which satisfies (2.1) up to terms of the order ϵ^3. Higher order approximations can be derived in a similar way, although the resulting equations rapidly become very complicated. Use of (2.2) in (2.1) gives

$$\{\phi_t A(\epsilon\mathbf{v}) + \phi_{\mathbf{x}_\alpha} B^{(\alpha)}(\epsilon\mathbf{v})\}\mathbf{v}_\theta + \epsilon\{A(\epsilon\mathbf{v})\mathbf{v}_t + B^{(\alpha)}(\epsilon\mathbf{v})\mathbf{v}_{\mathbf{x}_\alpha}\} = 0 \tag{2.5}$$

Here, $\mathbf{v}_\theta$ is the partial derivative of $\mathbf{v}$ at fixed $\mathbf{x}, t$ and $\mathbf{v}_t$, $\mathbf{v}_{\mathbf{x}_\alpha}$ are the partial derivatives at a fixed θ.

We note that if $\mathbf{v}(\mathbf{x}, t, \theta, \epsilon)$ satisfies (2.5) when $\theta = \epsilon^{-1}\phi(\mathbf{x}, t, \epsilon)$ (rather than for all θ, as is usually assumed in the method of multiple scales), then (2.2) gives a solution of the original equation (2.1). We are therefore free to regard any coefficient in (2.5) which do not contain derivatives as functions of $\mathbf{x}, t$ with θ evaluated at $(\phi(\mathbf{x}, t, \epsilon))/\epsilon$. Using (2.3) in (2.5) and Taylor expanding the coefficient matrices, we obtain

$$\begin{aligned}&\{\phi_t A(\epsilon\mathbf{v}_0) + \phi_{x_\alpha} B^{(\alpha)}(\epsilon\mathbf{v}_0)\}\mathbf{v}_{0\theta} + \epsilon\{[\phi_t A(\epsilon\mathbf{v}_0) + \phi_{x_\alpha} B^{(\alpha)}(\epsilon\mathbf{v}_0)]\mathbf{v}'_\theta \\ &+ A(\epsilon\mathbf{v}_0)\mathbf{v}_{0t} + B^{(\alpha)}(\epsilon\mathbf{v}_0)\mathbf{v}_{0x_\alpha}\} + \epsilon^2\{[\phi_t(\nabla_{\mathbf{u}} A)(\epsilon\mathbf{v}_0) \cdot \mathbf{v}' \\ &+ \phi_{x_\alpha}(\nabla_{\mathbf{u}} B^{(\alpha)})(\epsilon\mathbf{v}_0) \cdot \mathbf{v}']\mathbf{v}_{0\theta} + A(\epsilon\mathbf{v}_0)\mathbf{v}'_t + B^{(\alpha)}(\epsilon\mathbf{v}_0)\mathbf{v}'_{x_\alpha}\} = O(\epsilon^3).\end{aligned} \tag{2.6}$$

As we remarked above $\mathbf{v}(\mathbf{x}, t, \theta, \epsilon)$ is only required to satisfy this equation when $\theta = \phi/\epsilon$. We can therefore evaluate all the coefficients at this value of θ to obtain the equation

$$
\begin{aligned}
&[\phi_t A_0 + \phi_{x_\alpha} B_0^{(\alpha)}]\mathbf{v}_{0\theta} + \epsilon\{(\phi_t A_0 + \phi_{x_\alpha} B_0^{(\alpha)})\mathbf{v}'_\theta + A_0\mathbf{v}_{0t} + B_0^{(\alpha)}\mathbf{v}_{0x_\alpha}\} \\
&\quad + \epsilon^2[\{\phi_t(\nabla_{\mathbf{u}}A)_0\cdot\mathbf{v}' + \phi_{x_\alpha}(\nabla_{\mathbf{u}}B^{(\alpha)})_0\cdot\mathbf{v}'\}\mathbf{v}_{0\theta} + A_0\mathbf{v}'_t + B_0^{(\alpha)}\mathbf{v}'_{x_\alpha}] \\
&\qquad = O(\epsilon^3), \qquad (2.7)
\end{aligned}
$$

where the subscript 0 indicates that the coefficients are evaluated at $\mathbf{u} = \epsilon\mathbf{v}_0(\mathbf{x}, t, \epsilon^{-1}\phi(t, \mathbf{x}, \epsilon), \epsilon)$ so that they are functions of $\mathbf{x}, t$ and ϵ. For example,

$$
B_0^{(\alpha)}(\mathbf{x}, t, \epsilon) = B^{(\alpha)}(\epsilon\mathbf{v}_0(\mathbf{x}, t, \epsilon^{-1}\phi(\mathbf{x}, t, \epsilon), \epsilon)). \qquad (2.8)
$$

The three terms in (2.7) are not completely separated as coefficients of the powers of ϵ^0, ϵ and ϵ^2 are also dependent on ϵ. The first term, which is of order 1, vanishes up to this order and, therefore, we impose that it is exactly zero i.e.

$$
\{\phi_t A_0 + \phi_{x_\alpha} B_0^{(\alpha)}\}\mathbf{v}_{0\theta} = 0. \qquad (2.9)
$$

When we choose the leading term $\mathbf{v}_0$ in the high frequency asymptotic limit $\epsilon \to 0$ to satisfy this equation, the first term in (2.7) vanishes and we get a relation

$$
\begin{aligned}
&\{\phi_t A_0 + \phi_{x_\alpha} B_0^{(\alpha)}\}\mathbf{v}'_\theta + A_0\mathbf{v}_{0t} + B_0^{(\alpha)}\mathbf{v}_{0x_\alpha} \\
&\quad + \epsilon\{[\phi_t(\nabla_{\mathbf{u}}A)_0\cdot\mathbf{v}' + \phi_{x_\alpha}(\nabla_{\mathbf{u}}B^{(\alpha)})_0\cdot\mathbf{v}']\mathbf{v}_{0\theta} + A_0\mathbf{v}'_t + B_0^{(\alpha)}\mathbf{v}'_{x_\alpha}\} \\
&\qquad = O(\epsilon^2) \qquad (2.10)
\end{aligned}
$$

between $\mathbf{v}_0$ and $\mathbf{v}'$ with error of the order ϵ^2. To obtain a nontrivial solution for $\mathbf{v}_0$, we then require that ϕ satisfies the eikonal equation

$$
\det[\phi_t A_0(\mathbf{x}, t, \epsilon) + \phi_{x_\alpha}(\mathbf{x}, t, \epsilon)B_0^{(\alpha)}(\mathbf{x}, t, \epsilon)] = 0. \qquad (2.11)
$$

We note that this eikonal equation is associated with the function $\mathbf{u} = \epsilon\mathbf{v}_0(\mathbf{x}, t, \epsilon^{-1}\phi(\mathbf{x}, t, \epsilon), \epsilon)$ and thus we are able to incorporate leading order wave amplitude correction in the eikonal equation itself. We denote left and right null vectors associated with the phase $\phi(t, \mathbf{x}, \epsilon)$ and the perturbed state $u = \epsilon\mathbf{v}_0$ by $\mathbf{l}_0(\mathbf{x}, t, \epsilon)$ and $\mathbf{r}_0(\mathbf{x}, t, \epsilon)$, respectively, i.e $\mathbf{l}_0$ and $\mathbf{r}_0$ satisfy

$$
\mathbf{l}_0.(\phi_t A_0 + \phi_{x_\alpha} B_0^{(\alpha)}) = 0 \qquad (2.12)
$$

and

$$
(\phi_t A_0 + \phi_{x_\alpha} B_0^{(\alpha)})\mathbf{r}_0 = 0. \qquad (2.13)
$$

Here

$$
\mathbf{l}_0(\mathbf{x}, t, \epsilon) = \mathbf{l}(\mathbf{n}(\mathbf{x}, t, \epsilon), \epsilon\mathbf{v}_0), \qquad (2.14)
$$

$$
\mathbf{r}_0(\mathbf{x}, t, \epsilon) = \mathbf{r}(\mathbf{n}(\mathbf{x}, t, \epsilon), \epsilon\mathbf{v}_0), \qquad (2.15)
$$

where

$$
\mathbf{n}(\mathbf{x}, t, \epsilon) = \frac{\nabla\phi}{|\nabla\phi|}, \qquad \nabla\phi = (\phi_{x_1}, \phi_{x_2}, \ldots, \phi_{x_m}). \qquad (2.16)
$$

Also we normalize $\mathbf{l}_0$ so that

$$\mathbf{l}_0 A_0 \mathbf{r}_0 = 1. \tag{2.17}$$

A solution of (2.9) is given by

$$\mathbf{v}_0(\mathbf{x},t,\theta,\epsilon) = w(\mathbf{x},t,\theta,\epsilon)\mathbf{r}_0(\mathbf{x},t,\epsilon), \tag{2.18}$$

where w is an arbitrary scalar valued amplitude function. Taking the scalar product of (2.10) with the left null vector $\mathbf{l}_0$ we obtain

$$\begin{aligned}&\mathbf{l}_0(A_0\mathbf{v}_{0t} + B_0^{(\alpha)}\mathbf{v}_{0x_\alpha})\\ &\quad + \epsilon\mathbf{l}_0\{[\phi_t(\nabla_\mathbf{u}A)_0\cdot\mathbf{v}' + \phi_{x_\alpha}(\nabla_\mathbf{u}B^{(\alpha)})_0\cdot\mathbf{v}']\mathbf{v}_{0\theta} + A_0\mathbf{v}'_t + B_0^{(\alpha)}\mathbf{v}'_{x_\alpha}\}\\ &\qquad = O(\epsilon^2).\end{aligned} \tag{2.19}$$

To eliminate $\mathbf{v}'$ from this equation, we solve (2.10) iteratively for $\mathbf{v}'$ in terms of $\mathbf{v}_0$. In order that the eliminant has an error of order ϵ^2 consistent with (2.19), we note that we need to solve $\mathbf{v}'$ with error of order ϵ i.e we consider only the leading order terms in (2.10)

$$\{\phi_t A_0 + \phi_{x_\alpha}B_0^{(\alpha)}\}\mathbf{v}'_\theta + A_0\mathbf{v}_{0t} + B_0^{(\alpha)}\mathbf{v}_{0x_\alpha} = O(\epsilon). \tag{2.20}$$

We use (2.18) in (2.20). A solution of the resulting equation for $\mathbf{v}'$ is then

$$\mathbf{v}'(\mathbf{x},t,\theta,\epsilon) = b_t(\mathbf{x},t,\theta,\epsilon)\mathbf{s}'_0 + b_{x_\beta}(\mathbf{x},t,\theta,\epsilon)\mathbf{s}_0^{(\beta)} + b(\mathbf{x},t,\theta,\epsilon)\mathbf{s}_0 + O(\epsilon), \tag{2.21}$$

where b is the scalar amplitude such that

$$b_\theta = w \tag{2.22}$$

and the vectors $\mathbf{s}_0(\mathbf{x},t,\epsilon)$, $\mathbf{s}'_0(t,\mathbf{x},\epsilon)$ and $\mathbf{s}_0{}^{(\beta)}(\mathbf{x},t,\epsilon)$ satisfy

$$\begin{aligned}(\phi_t A_0 + \phi_{x_\alpha}B_0^{(\alpha)})\mathbf{s}_0 &= -(A_0\mathbf{r}_{0t} + B_0^{(\alpha)}\mathbf{r}_{0x_\alpha})\\ &\quad + (\mathbf{l}_0(A_0\mathbf{r}_{0t} + B_0^{(\alpha)}\mathbf{r}_{0x_\alpha}))A_0\mathbf{r}_0,\end{aligned} \tag{2.23a}$$

$$(\phi_t A_0 + \phi_{x_\alpha}B_0^{(\alpha)})\mathbf{s}'_0 = -(A_0\mathbf{r}_0) + (\mathbf{l}_0 A_0\mathbf{r}_0)A_0\mathbf{r}_0, \tag{2.23b}$$

$$(\phi_t A_0 + \phi_{x_\alpha}B_0^{(\alpha)})\mathbf{s}_0^\beta = -(B_0^{(\beta)}\mathbf{r}_0) + (\mathbf{l}_0 B_0^{(\beta)}\mathbf{r}_0)A_0\mathbf{r}_0. \tag{2.24}$$

These equations do not have a unique solution. This is because there is some arbitrariness in how $\mathbf{v}$ is decomposed into $\mathbf{v}_0$ and $\mathbf{v}'$. But if we impose the condition (2.4) on $\mathbf{v}'$, then we choose the unique solutions of (2.23)–(2.24) such that

$$\mathbf{l}_0\mathbf{s}_0 = \mathbf{l}_0\mathbf{s}'_0 = \mathbf{l}_0\mathbf{s}_0{}^\beta = 0. \tag{2.25}$$

Finally, use of (2.18) and (2.21) in (2.19) gives the following transport equation for w,

$$\begin{aligned}&w_t + \chi_{\alpha_0}w_{x_\alpha} - \Omega w + \epsilon[(\Gamma^t b_t + \Gamma^\alpha b_{x_\alpha} + \Gamma b)w_\theta\\ &\qquad + Wb_t + V^\alpha b_{x_\alpha} + D^{\alpha\beta} + Eb] = O(\epsilon^2).\end{aligned} \tag{2.26}$$

Note that $D^{\alpha\beta}$ contains linear terms in the second order derivatives of b as seen below. The coefficients are functions of $(\mathbf{x},t,\epsilon)$ given by

$$\chi_{\alpha_0} = \mathbf{l}_0 B_0^{(\alpha)}\mathbf{r}_0,$$

$$\Omega = -(\mathbf{l}_0 A_0\mathbf{r}_{0t} + \mathbf{l}_0 B_0^{(\alpha)}\mathbf{r}_{0x_\alpha}),$$

$$\begin{aligned}
\Gamma &= \mathbf{l}_0\left(\phi_t(\nabla_{\mathbf{u}}A)_0 + \phi_{x_\alpha}(\nabla_{\mathbf{u}}B^{(\alpha)})_0\right)\cdot \mathbf{s}_0\,\mathbf{r}_0,\\
\Gamma^t &= \mathbf{l}_0\left(\phi_t(\nabla_{\mathbf{u}}A)_0 + \phi_{x_\beta}(\nabla_{\mathbf{u}}B^{(\beta)})_0\right)\cdot \mathbf{s}_0'\,\mathbf{r}_0,\\
\Gamma^\alpha &= \mathbf{l}_0\left(\phi_t(\nabla_{\mathbf{u}}A)_0 + \phi_{x_\beta}(\nabla_{\mathbf{u}}B^{(\beta)})_0\right)\cdot \mathbf{s}_0^{(\alpha)}\,\mathbf{r}_0,\\
W &= \mathbf{l}_0\left(A_0\mathbf{s}_0 + A_0\mathbf{s}_{0t}' + B_0^{(\beta)}\mathbf{s}_{0x_\beta}'\right),\\
V^\alpha &= \mathbf{l}_0\left(B_0^{(\alpha)}\mathbf{s}_0 + A_0\mathbf{s}_{0t}^{(\alpha)} + B_0^{(\beta)}\mathbf{s}_{0x_\beta}^{(\alpha)}\right),\\
D^{\alpha\beta} &= \mathbf{l}_0\{A_0\mathbf{s}_0' b_{tt} + A_0\mathbf{s}_0^{(\beta)}b_{x_\beta t} + B_0^{(\alpha)}\mathbf{s}_0' b_{tx_\alpha} + B_0^{(\alpha)}\mathbf{s}_0^{(\beta)}b_{x_\alpha x_\beta}\},\\
E &= \mathbf{l}_0(A_0\mathbf{s}_{0t} + B_0^{(\alpha)}\mathbf{s}_{0x_\alpha}).
\end{aligned}\tag{2.27}$$

3. Ray formulation of the asymptotic equations

The eikonal equation (2.9) can be equivalently written in the form

$$Q \equiv \phi_t(\mathbf{l}_0A_0\mathbf{r}_0) + \phi_{x_\alpha}(\mathbf{l}_0B_0^{(\alpha)}\mathbf{r}_0) = 0, \quad \alpha = 1,\dots,m. \tag{3.1}$$

From the characteristic equations of (3.1) we obtain

$$\frac{\mathrm{d}x_\alpha}{\mathrm{d}s} = \frac{\partial Q}{\partial\phi_{x^\alpha}} = \mathbf{l}_0B_0^{(\alpha)}\mathbf{r}_0 = \chi_{\alpha 0}, \tag{3.2}$$

$$\frac{\mathrm{d}t}{\mathrm{d}s} = \frac{\partial Q}{\partial\phi_t} = \mathbf{l}_0A_0\mathbf{r}_0 = 1, \tag{3.3}$$

$$\frac{\mathrm{d}\phi_{x_\alpha}}{\mathrm{d}s} = -\frac{\partial Q}{\partial x^\alpha} = \phi_t(\mathbf{l}_0A_{0x_\alpha}\mathbf{r}_0) + \phi_{x_\gamma}(\mathbf{l}_0B_{0x^\alpha}^{(\gamma)}\mathbf{r}_0), \quad \gamma = 1,\dots,m. \tag{3.4}$$

Now for a fixed t, $\phi(\mathbf{x},t,\epsilon) = 0$ represents a wavefront in $\mathbf{x}$-space with unit normal

$$\mathbf{n} = \frac{\nabla\phi}{|\nabla\phi|}, \quad \nabla = (\partial_{x_1},\dots,\partial_{x_m}).$$

The differential equation for $\mathbf{n}$ is

$$\frac{\mathrm{d}n_\alpha}{\mathrm{d}s} = -n_\beta\mathbf{l}_0\left(-c_0\frac{\partial A_0}{\partial\eta_\beta^\alpha} + n_\gamma\frac{\partial B_0^\gamma}{\partial\eta_\beta^\alpha}\right)\mathbf{r}_0 \equiv \Psi_{\alpha 0}, \quad \text{say} \tag{3.5}$$

where

$$\frac{\partial}{\partial\eta_\beta^\alpha} = n_\beta\frac{\partial}{\partial x_\alpha} - n_\alpha\frac{\partial}{\partial x^\beta}, \quad \beta = 1,2,\dots,m \tag{3.6}$$

and

$$c_0 = \frac{\phi_t}{|\nabla\phi|}. \tag{3.7}$$

The operator

$$\frac{\mathrm{d}}{\mathrm{d}s} = \frac{\partial}{\partial t} + \chi_{\alpha 0}\frac{\partial}{\partial x_\alpha} \tag{3.8}$$

appearing on the left hand side of (3.2)–(3.4) and (3.5), and $\partial/\partial\eta_\beta^\alpha$ defined above are in direction tangential to a characteristic surface $\phi(\mathbf{x},t) =$ constant in $(\mathbf{x},t)$ space. In addition the derivatives $\partial/\partial\eta^{\alpha}_{\beta}$ are tangential to a wavefront $\phi(\mathbf{x},t) = \text{constant}$ with $t =$ constant in $(x_1,\dots,x_m)$-space. Because of the choice (2.17), $\mathrm{d}/\mathrm{d}s$ represents time rate of change along a ray and may be denoted by the symbol $\mathrm{d}/\mathrm{d}t$. The transport equation (2.26) can then be written as

$$\frac{\mathrm{d}w}{\mathrm{d}s} = \Omega w - \epsilon[(\Gamma^t b_t + \Gamma^{\alpha} b_{x_\alpha} + \Gamma b)w_\theta + W b_t + V^{\alpha} b_{x_\alpha} + D^{\alpha\beta} + Eb] + O(\epsilon^2). \tag{3.9}$$

The equations (3.2), (3.5) and (3.9) form a complete set of equations of the nonlinear ray theory with error $O(\epsilon^2)$. The amplitude $\mathbf{u} = \epsilon\mathbf{v}_0 = \epsilon w\mathbf{r}_0$ up to first order in ϵ appears in the bicharacteristic velocity $\chi_{\alpha_0}(\mathbf{x},t,\epsilon)$ and the rate of turning Ψ_{α_0} of the rays, given in a complicated way.

The interesting and important point of this weakly nonlinear ray theory (WNLRT) is that the transport equation (3.9) for w along nonlinear rays is coupled to the ray equations (3.2) and (3.5), which correspond to the leading order wave amplitude w. Earlier transport equation, derived by Prasad [12] for a general hyperbolic system*, were along the exact nonlinear rays corresponding to the exact solution $\mathbf{u}$ in the form (2.2). Prasad derived the transport equation on geometrical consideration by approximating the system (2.1) in the neighbourhood of the exact characteristic surface in space-time. Proper interpretation of transport equation along the nonlinear rays corresponding to leading order amplitude w has lead to physically realistic solutions [19, 13, 14, 18].

To make these equations more tractable, we approximate $\mathbf{l}_0$ and $\mathbf{r}_0$ defined by (2.12) as follows. We now define $\bar{\mathbf{l}}$ and $\bar{\mathbf{r}}$ as

$$\bar{\mathbf{l}} = \mathbf{l}(\mathbf{n},0) \quad \bar{\mathbf{r}} = \mathbf{r}(\mathbf{n},0).$$

Then

$$\begin{aligned} \mathbf{l}_0 &= \bar{\mathbf{l}} + \epsilon\{(\nabla_{\mathbf{u}}\mathbf{l})_0 \cdot \mathbf{v}_0\} + O(\epsilon^2), \\ &= \bar{\mathbf{l}} + \epsilon\{(\nabla_{\mathbf{u}}\mathbf{l})_0 \cdot \mathbf{r}_0\}w + O(\epsilon^2), \end{aligned} \tag{3.10}$$

where $(\nabla_{\mathbf{u}}\mathbf{l})_0$ is the value of $(\nabla_{\mathbf{u}}\mathbf{l})$ evaluated at $\mathbf{u} = 0$ keeping $\mathbf{n}$ fixed and is a notation different from that introduced by eq. (2.8) for the use of the subscript 0. Similarly

$$\mathbf{r}_0 = \bar{\mathbf{r}} + \epsilon\{(\nabla_{\mathbf{u}}\mathbf{r})_0 \cdot \mathbf{r}_0\}w + O(\epsilon^2). \tag{3.11}$$

The vectors $\bar{\mathbf{l}}$ and $\bar{\mathbf{r}}$ still depend on the leading order term $\mathbf{v}_0$ in the solution and the nonlinear phase ϕ, through $\mathbf{n}$. Also if

$$A_* = A(\mathbf{u} = 0) \quad \text{and} \quad B^{(\alpha)}_* = B^{(\alpha)}(\mathbf{u} = 0) \quad \text{are constant matrices,} \tag{3.12}$$

then we have

$$A_0 = A_* + \epsilon(\nabla_u A)_* \cdot \mathbf{r}_0 w + O(\epsilon^2) \tag{3.13}$$

and

$$B^{(\alpha)}_0 = B^{(\alpha)}_* + \epsilon(\nabla_u B^{(\alpha)})_* \cdot \mathbf{r}_0 w + O(\epsilon^2), \tag{3.14}$$

*This was inspired by the work of K.E. Gubkin in 1958 for gasdynamic equations (see *PMM J. Appl. Math. Mech.* **22** 787–793) where $(\nabla_{\mathbf{u}} B^{(\alpha)})_*$ is the value of $(\nabla_{\mathbf{u}} B^{(\alpha)})$ evaluated at $\mathbf{u} = 0$. The important point in simplifying the equations now is to realise that a nonlinear wavefront given by the phase function $\phi(\mathbf{x}, t, \epsilon)$ may differ significantly from the corresponding linear wavefront given by the linear phase function $\phi^*(\mathbf{x}, t)$. This can be seen from the large number of results we have presented in the earlier papers including that by Prasad and Sangeeta [18]. The partial derivatives ϕ_{x_α} of the nonlinear phase and $\phi^*_{x_\alpha}$ of the linear phase (i.e the unit normal $\mathbf{n}$ of a nonlinear wavefront and $\mathbf{n}_*$ of the corresponding linear wavefront) also differ significantly. One may think that the nonlinear ray theory which is being considered here may be valid only on the length scale over which the linear theory or Choquet–Bruhat's nonlinear theory are valid. But this is not so. In the derivation of this theory we have made no reference to the length scales associated with the linear theory. The numerical results of Prasad and Sangeeta [18] show that this theory is valid even in a caustic region where the normal $\mathbf{n}$ of a nonlinear wavefront and $\mathbf{n}_*$ of the corresponding linear wavefront differ very much. In fact, the theory is valid on much larger length scale than the radii of curvature of the initial wavefront. Therefore, while trying to make further approximation in some of the terms in (3.2), (3.3), (3.5) and (3.9) we keep $\mathbf{n}$ and the operators $\partial/\partial\eta^\alpha_\beta$ (tangential derivatives on the nonlinear wavefront) unchanged and use Taylor's expansion with respect to $\epsilon\mathbf{v_0}$ at $\mathbf{0}$. Following this we can approximate some of the terms as follows

$$\begin{aligned}\mathbf{l}_0 B_0^{(\alpha)} \mathbf{r}_0 = {} & \bar{\mathbf{l}} B_*^{(\alpha)} \bar{\mathbf{r}} + \epsilon[(\nabla_{\mathbf{u}} \mathbf{l})_0 \cdot \bar{\mathbf{r}} B_*^{(\alpha)} \bar{\mathbf{r}} + \bar{\mathbf{l}}((\nabla_{\mathbf{u}} B^{(\alpha)})_* \cdot \bar{\mathbf{r}}) \bar{\mathbf{r}} \\ & + \bar{\mathbf{l}} B_*^{(\alpha)} (\nabla_{\mathbf{u}} \mathbf{r})_0 \cdot \bar{\mathbf{r}}] w + O(\epsilon^2), \end{aligned} \tag{3.15}$$

$$\begin{aligned}\mathbf{l}_0 A_0 \mathbf{r}_{0t} = {} & \bar{\mathbf{l}} A_* \bar{\mathbf{r}}_t + \epsilon[(\nabla_{\mathbf{u}} \mathbf{l})_0 \cdot \bar{\mathbf{r}} A_* \bar{\mathbf{r}}_t + \bar{\mathbf{l}}((\nabla_{\mathbf{u}} A)_* \cdot \bar{\mathbf{r}}) \bar{\mathbf{r}}_t + \bar{\mathbf{l}} A_* (\nabla_{\mathbf{u}} \mathbf{r})_{0t} \cdot \bar{\mathbf{r}} \\ & + \bar{\mathbf{l}} A_* (\nabla_{\mathbf{u}} \mathbf{r})_0 \cdot \bar{\mathbf{r}}_t] w + \epsilon \bar{\mathbf{l}} A_* (\nabla_{\mathbf{u}} \mathbf{r})_0 \cdot \bar{\mathbf{r}} w_t + O(\epsilon^2). \end{aligned} \tag{3.16}$$

and

$$\begin{aligned}\mathbf{l}_0 B_0 \mathbf{r}_{0_{x_\alpha}} = {} & \bar{\mathbf{l}} B_*^{(\alpha)} \bar{\mathbf{r}} + \epsilon[(\nabla_{\mathbf{u}} \mathbf{l})_0 \cdot \bar{\mathbf{r}} B_*^{(\alpha)} \bar{\mathbf{r}}_{x_\alpha} + \bar{\mathbf{l}}((\nabla_{\mathbf{u}} B^{(\alpha)})_* \cdot \bar{\mathbf{r}}) \bar{\mathbf{r}}_{x_\alpha} + \bar{\mathbf{l}} A_* (\nabla_{\mathbf{u}} \bar{\mathbf{r}})_{0x_\alpha} \cdot \bar{\mathbf{r}} \\ & + \bar{\mathbf{l}} B_*^{(\alpha)} (\nabla_{\mathbf{u}} \mathbf{r})_0 \cdot \bar{\mathbf{r}}_{x_\alpha}] w + \epsilon \bar{\mathbf{l}} B_*^{(\alpha)} (\nabla_{\mathbf{u}} \mathbf{r})_0 \cdot \bar{\mathbf{r}} w_{x_\alpha} + O(\epsilon^2). \end{aligned} \tag{3.17}$$

Therefore

$$\mathbf{l}_0 A_0 \mathbf{r}_{0t} + \mathbf{l}_0 B_0^{(\alpha)} \mathbf{r}_{0x_\alpha} = \bar{\mathbf{l}} A_* \bar{\mathbf{r}}_t + \bar{\mathbf{l}} B_*^{(\alpha)} \bar{\mathbf{r}}_{x_\alpha} + O(\epsilon) = -\bar{\Omega} + O(\epsilon), \tag{3.18}$$

where

$$\bar{\Omega} = -(\bar{\mathbf{l}} A_* \mathbf{r}_t + \bar{\mathbf{l}} B_*^{(\alpha)} \bar{\mathbf{r}}_{x_\alpha}). \tag{3.19}$$

Substituting (3.15) to (3.19) in (3.2), (3.5) and (3.9) and retaining terms only up to order ϵ we get the full set of equations of WNLRT (note $\mathrm{d}/\mathrm{d}s = \mathrm{d}/\mathrm{d}t$)

$$\begin{aligned}\frac{\mathrm{d}x_\alpha}{\mathrm{d}t} = {} & \bar{\mathbf{l}} B_*^\alpha \bar{\mathbf{r}} + \epsilon[(\nabla_{\mathbf{u}} \mathbf{l})_0 \cdot \bar{\mathbf{r}} B_*^{(\alpha)} \bar{\mathbf{r}} + \bar{\mathbf{l}}((\nabla_{\mathbf{u}} B^{(\alpha)})_* \cdot \bar{\mathbf{r}}) \bar{\mathbf{r}} \\ & + \bar{\mathbf{l}} B_*^{(\alpha)} (\nabla_{\mathbf{u}} \mathbf{r})_0 \bar{\mathbf{r}}] w + O(\epsilon^2), \end{aligned} \tag{3.20}$$

$$\begin{aligned}\frac{\mathrm{d}n_0^\alpha}{\mathrm{d}t} = {} & -\epsilon \mathbf{n}^\beta \bar{\mathbf{l}} \left[\left\{ -\bar{c}(\nabla_{\mathbf{u}} A)_* \frac{\partial \bar{\mathbf{r}}}{\partial \eta_\beta^\alpha} + n_\gamma (\nabla_{\mathbf{u}} B^\gamma)_* \frac{\partial \bar{\mathbf{r}}}{\partial \eta_\beta^\alpha} \right\} w \right. \\ & \left. + \{-\bar{c}(\nabla_{\mathbf{u}} A)_* \bar{\mathbf{r}} + n_\gamma (\nabla_{\mathbf{u}} B^\gamma)_* \bar{\mathbf{r}}\} \frac{\partial w}{\partial \eta_\beta^\alpha} \right] \bar{\mathbf{r}} + O(\epsilon^2), \\ & \beta = 1, 2 \ldots, m \end{aligned} \tag{3.21}$$

where

$$\bar{c} = c_0(\mathbf{n}, \mathbf{u} = 0)$$

and we note that $\psi_*^{(\alpha)}$ is zero because A_* and $B_*^{(\alpha)}$ are constants and

$$\begin{aligned}\frac{\mathrm{d}w}{\mathrm{d}t} = {} & \bar{\Omega} w + \epsilon[(\nabla_{\mathbf{u}} l)_0 \cdot \bar{\mathbf{r}} A_* \bar{\mathbf{r}}_t + \bar{\mathbf{l}}((\nabla_{\mathbf{u}} A)_* \cdot \bar{\mathbf{r}})\bar{\mathbf{r}}_t + \bar{\mathbf{l}} A_* (\nabla_{\mathbf{u}} \mathbf{r})_0 \cdot \bar{\mathbf{r}}_t] w \\ & + \epsilon[(\nabla_{\mathbf{u}} \mathbf{l})_0 \cdot \bar{\mathbf{r}} B_*^{(\alpha)} \bar{\mathbf{r}}_{x_\alpha} + \bar{\mathbf{l}}((\nabla_{\mathbf{u}} B^{(\alpha)})_* \cdot \bar{\mathbf{r}})\bar{\mathbf{r}}_{x_\alpha} + \bar{\mathbf{l}} B_*^{(\alpha)} (\nabla_{\mathbf{u}} \mathbf{r})_0 \cdot \bar{\mathbf{r}}_{x_\alpha}] w \\ & + \epsilon\{\bar{\mathbf{l}} A_* (\nabla_{\mathbf{u}} \mathbf{r})_0 \cdot \bar{\mathbf{r}} w_t + \bar{\mathbf{l}} B_*^{(\alpha)} (\nabla_{\mathbf{u}} \mathbf{r})_0 \cdot \bar{\mathbf{r}} w_{x_\alpha}\} \\ & - \epsilon[(\Gamma^t b_t + \Gamma^\alpha b_{x_\alpha} + \Gamma b) w_\theta + W b_t + V^\alpha b_{x_\alpha} + D^{\alpha\beta} + E b] \\ & + O(\epsilon^2). \end{aligned} \tag{3.22}$$

If the terms of the order ϵ are also neglected in the ray equations (3.20) and (3.21), these equations decouple from the transport equation (3.22) and give the linear rays. In order to retain the nonlinear effects it is necessary to retain in the ray equations, terms atleast up to order ϵ. The situation for the transport equation (3.22) is different. Exact solution [19] and numerical results [18] show that inclusion of order ϵ terms in (3.20) and (3.21) changes $\bar{\Omega}$ by order 1 in the caustic region leading to order 1 change in the value of w in finite time. This is in contrast to what we expect in a perturbation method. But it is not surprising when we note that the neglect of $O(\epsilon)$ terms (3.20) and (3.21) (i.e. linear theory) changes $\bar{\Omega}$ from a finite curvature to infinite curvature in the caustic region which is reached in finite time. It is different with the transport equation (3.22) which with only the first term on the right hand side always leads to a finite value of w everywhere. During the competition of convergence of linear rays and opposing effect of nonlinearity, a balance is reached which leads to a finite change in $\bar{\Omega}$. There is no mathematical proof so far for the amplitude to be finite due to nonlinearity but extensive numerical computation with small (but not very small) values of amplitude w leads to this conjecture. In all these cases the effect of inclusion of the terms of order ϵ in (3.22) will remain small in finite time. As stated in the abstract and the end of the introduction, we have indeed deduced a weakly nonlinear theory (i.e. eqs (3.20)–(3.22)) in which w has error $O(\epsilon^2)$ (i.e. the solution $\mathbf{u}$ has error $O(\epsilon^3)$). However, in the solution of the simpler WNLRT (i.e. eqs (3.20), (3.21) and (3.23)) the amplitude w has error $O(\epsilon)$. Thus, to get only the leading order correction to the amplitude, it is not necessary to retain the last four terms in (3.22) which are multiplied by ϵ and then we get

$$\frac{\mathrm{d}w}{\mathrm{d}t} = \bar{\Omega} w. \tag{3.23}$$

This transport equation looks exactly the same as the linear transport equation but it contains now all leading order nonlinear effects since in it $\mathrm{d}w/\mathrm{d}t$ represents time rate of change along the nonlinear rays and $\mathbf{n}$ appearing in $\bar{\Omega}$ is the normal of the nonlinear wavefront. In fact the equation (3.23) along with the equations (3.20) and (3.21) is equivalent to the transport equation

$$\begin{aligned} w_t + \{ & \bar{\mathbf{l}} B_*^{(\alpha)} \bar{\mathbf{r}} + \epsilon[(\nabla_{\mathbf{u}} \mathbf{l})_0 \cdot \bar{\mathbf{r}} B_*^{(\alpha)} \bar{\mathbf{r}} + \bar{\mathbf{l}}((\nabla_{\mathbf{u}} B^{(\alpha)})_* \cdot \bar{\mathbf{r}})\bar{\mathbf{r}} \\ & + \bar{\mathbf{l}} B_*^{(\alpha)} (\nabla_{\mathbf{u}} \mathbf{r})_0 \cdot \bar{\mathbf{r}}] w\} w_{x_\alpha} = \bar{\Omega} w \end{aligned} \tag{3.24}$$

and $\bar{\Omega}$, which contains derivatives of $\mathbf{n}$, remains finite everywhere including the points on the caustic, where the corresponding value Ω^* by linear theory tends to infinity. The equations (3.20), (3.21) and (3.23) form a coupled system of equations of a nonlinear ray theory. Retaining the other terms of order ϵ in (3.22) will modify the results only by effects of order ϵ^2 since the neglected terms are actually of order ϵ^2 in the original equation (2.6). Equations (3.20), (3.21) and (3.23) are exactly the same as the equations obtained for a nonlinear ray theory by Prasad [15] (see also [12, 19, 14]). In these earlier papers w is of order ϵ, i.e w there is same as ϵw here.

4. Comparison with other theories

The WNLRT developed in the last two sections is valid over a length scale L over which the assumptions involved in the derivation of the equations are valid. This length L can be determined only from the solution of this approximate theory. One exact solution, called composite simple wave solution in Ravindran and Prasad [19] and Prasad [14], and extensive numerical solution by Prasad and Sangeeta [18] show that this L is large compared to the length scale R of the order of principal radii of curvature of the initial wavefront. The Choquet–Bruhat's nonlinear theory is valid over a length scale L_c which is small compared to R. On this scale L_c, the linear and nonlinear wavefronts are not only close but have same shape and the amplitude given by the linear theory remains small. Thus $L_c/R \ll 1 \ll R/L$. We shall show that over the length scale L_c, the equation (3.23) reduces to the leading order equation obtained from Choquet–Bruhat's theory in addition to some extra terms which can be neglected. We examine the (3.23) over a length scale L_c. On this length scale, the linear wavefront and the corresponding nonlinear wavefront originating from a same initial wavefront are close to one another and their unit normals denoted respectively by $\mathbf{n}_*$ and $\mathbf{n}$ differ by a quantity of order ϵ. We denote the rate of change along the linear ray by $\mathrm{d}^*/\mathrm{d}s^*$ i.e.

$$\frac{\mathrm{d}^*}{\mathrm{d}s^*} = \mathbf{l}_* A_*^\alpha \mathbf{r}_* \frac{\partial}{\partial x^\alpha}, \quad (x_\alpha) = (x_0 = t, x_1 = x, \dots x_m = x_m), \tag{4.1}$$

where we have not set $\mathbf{l}A\mathbf{r} = 1$, and used $A^0 = A$, $A^\alpha = B^{(\alpha)}$ and

$$\mathbf{l}_* = \bar{\mathbf{l}}(\mathbf{n}_*, 0), \mathbf{r}_* = \bar{\mathbf{r}}(\mathbf{n}_*, 0). \tag{4.2}$$

The summation convention in this section extends on the range $0, 1, 2, \dots, m$. The rate of change $\mathrm{d}/\mathrm{d}s$ along the nonlinear ray (see eqs (3.23) and (3.24) for $|\mathbf{n} - \mathbf{n}_*| = O(\epsilon)$) can be written as

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s} = {} & \mathbf{l}_* A_*^\alpha \mathbf{r}_* \frac{\partial}{\partial x_\alpha} + \epsilon \Big\{ \Big((\nabla_\mathbf{n} \bar{\mathbf{l}})_* \cdot \Big(\frac{\mathbf{n} - \mathbf{n}_*}{\epsilon} \Big) \Big) A_*^\alpha \mathbf{r}_* \\ & + \mathbf{l}_* A_*^\alpha \Big((\nabla_\mathbf{n} \bar{\mathbf{r}})_* \cdot \Big(\frac{\mathbf{n} - \mathbf{n}_*}{\epsilon} \Big) \Big) \Big\} \frac{\partial}{\partial x_\alpha} + \epsilon w [((\nabla_\mathbf{u} \mathbf{l})_* \cdot \mathbf{r}_*) A_*^\alpha \mathbf{r}_* \\ & + \mathbf{l}_* ((\nabla_\mathbf{u} A^\alpha)_* \cdot \mathbf{r}_*) \mathbf{r}_* + \mathbf{l}_* A_*^\alpha ((\nabla_\mathbf{u} \mathbf{r})_* \cdot \mathbf{r}_*)] \frac{\partial}{\partial x_\alpha} + 0(\epsilon^2),\end{aligned} \tag{4.3}$$

where $\nabla_\mathbf{n} = \left(\frac{\partial}{\partial n_1}, \dots, \frac{\partial}{\partial n_n} \right)$. The middle term in the square bracket is important and we write it along with the first term on the right hand side of (4.3). Thus

$$\frac{\mathrm{d}}{\mathrm{d}s} = \frac{\mathrm{d}^*}{\mathrm{d}s^*} + \epsilon \{ \mathbf{l}_* (\nabla_\mathbf{u} A^\alpha)_* \mathbf{r}_*) \mathbf{r}_* \} w \frac{\partial}{\partial x_\alpha} + \epsilon w S^\alpha \frac{\partial}{\partial x_\alpha} + \epsilon T^\alpha \frac{\partial}{\partial x_\alpha} + 0(\epsilon^2), \tag{4.4}$$

where

$$S^{\alpha} = ((\nabla_{\mathbf{u}}\mathbf{l})_* \cdot \mathbf{r}_*)A_*^{\alpha}\mathbf{r}_* + \mathbf{l}_* A_*^{\alpha}((\nabla_{\mathbf{u}}\mathbf{r})_* \cdot \mathbf{r}_*) \tag{4.5}$$

$$T^{\alpha} = \left((\nabla_{\mathbf{n}}\bar{\mathbf{l}})_* \cdot \left(\frac{\mathbf{n} - \mathbf{n}_*}{\epsilon}\right)\right) A_*^{\alpha}\mathbf{r}_* + \mathbf{l}_* A_*^{\alpha}\left((\nabla_{\mathbf{n}}\bar{\mathbf{r}})_* \cdot \left(\frac{\mathbf{n} - \mathbf{n}_*}{\epsilon}\right)\right). \tag{4.6}$$

The second term in (4.4) contains in it the nonlinear stretching of the rays as given in Choquet–Bruhat's theory. In fact, if we make a transformation from (x_α)-coordinates to $(\phi^*, y^1, \ldots, y^n)$-coordinates (where ϕ^* is the linear phase function)

$$\phi^* = \phi^*(x^0, x_1, \ldots, x_m), \quad y_\alpha = x_\alpha, \quad \alpha = 1, 2, \ldots n, \tag{4.7}$$

then

$$\frac{\partial}{\partial x^0} = \phi^*_{x^0}\frac{\partial}{\partial x^0}, \quad \frac{\partial}{\partial x_\alpha} = \phi^*_{x_\alpha}\frac{\partial}{\partial \phi^*} + \frac{\partial}{\partial y^\alpha}, \quad \alpha = 1, 2, \ldots, n \tag{4.8}$$

so that with $\theta^* = \frac{\phi^*}{\epsilon}$,

$$\epsilon\{\mathbf{l}_*((\nabla_u A^\alpha)_* \cdot \mathbf{r}_*)\mathbf{r}_*\}w\frac{\partial}{\partial x_\alpha} = \mathcal{G}w\frac{\partial}{\partial \theta^*} + 0(\epsilon),$$

where

$$\mathcal{G} = \{\mathbf{l}_*(\phi^*_{x_\alpha}(\nabla_u A^\alpha)_* \cdot \mathbf{r}_*)\mathbf{r}_*\} \tag{4.9}$$

since $\partial/\partial\theta^* = (1/\epsilon)\partial/\partial\phi^*$. Further

$$\epsilon S^\alpha \frac{\partial}{\partial x_\alpha} = \{((\nabla_u \mathbf{l})_* \cdot \mathbf{r}_*)(A_*^\alpha \phi^*_{x_\alpha}\mathbf{r}_*) + (\mathbf{l}_* A_*^\alpha \phi^*_{x_\alpha})((\nabla_u \mathbf{r})_* \mathbf{r}_*)\}\frac{\partial}{\partial \theta^*} + 0(\epsilon) \tag{4.10}$$

in which all terms of order one vanish because $A_*^\alpha\phi^*_{x_\alpha}\mathbf{r}_* = 0$, and $\mathbf{l}_* A_*^\alpha \phi^*_{x_\alpha} = 0$.

On the length scale L_c, $\mathbf{n} - \mathbf{n}_* = 0(\epsilon)$, so that

$$\epsilon T^\alpha \frac{\partial}{\partial x_\alpha} = T^\alpha \phi^*_{x_\alpha}\frac{\partial}{\partial \theta^*} + 0(\epsilon) \tag{4.11}$$

and here too all the terms of order one vanish due to the same reason i.e. $A_*^\alpha\phi^*_{x_\alpha}\mathbf{r}_* = 0$ and $\mathbf{l}_* A_*^\alpha \phi^*_{x_\alpha} = 0$. Thus, to the leading order, the transport equations (3.23) or (3.24) reduces to the Choquet–Bruhat's transport equation

$$\frac{\mathrm{d}^*}{\mathrm{d}t^*}w + \mathcal{G}ww_{\theta^*} + \Omega_* w = 0 \tag{4.12}$$

(see [5]). Note that the assumption $|\mathbf{n} - \mathbf{n}^*| = O(\epsilon)$ breaks down as soon the nonlinear wavefront starts approaching a caustic region of the linear theory.

One of the most interesting outcome of this theory is a derivation of the weak shock ray theory ([14], p. 95), from the WNLRT consisting of the eqs (3.20), (3.21) and (3.24). Shock ray theory consists of the shock ray equations, and an infinite system of compatibility conditions. Unlike the WNLRT, shock ray theory is exact because ϵ is of the order of the shock thickness which is zero in the inviscid theory and hence the high frequency approximation is exactly satisfied. But the shock ray theory is as difficult as the original problem, in fact more difficult due to horrendously long expressions present even in the first few (say the second itself) of the infinite number of compatibility conditions involved in it. Infinite number of equations remain involved even if weak shock assumption is made. As mentioned here, the weak shock ray theory can be derived from the WNLRT of this paper. This derivation is not only simple but also much more transparent for the Euler's equations of gas dynamics, which we shall do in the next section. In passing, we mention that an attempt has been made in showing such a relation between WNLRT and shock ray theory by Anile *et al* [1] pp. 85–87) without making any distinction between a linear, nonlinear and shock rays.

5. Nonlinear waves in a polytropic gas

Wave propagation in a gas (or in any continuous media) has as its foundation, the three basic conservation laws of physics: conservation of mass, momentum and energy. These laws of physics allow us to derive the *field equations* which for a polytropic gas are expressed in terms of the fluid velocity $\mathbf{q}$, density ρ and pressure p. The equations of motion are

$$\rho_t + \langle \nabla, \rho\mathbf{q} \rangle = 0, \tag{5.1a}$$

$$\mathbf{q}_t + \langle \mathbf{q}, \nabla \rangle \mathbf{q} + \frac{1}{\rho} \nabla p = 0 \tag{5.1b}$$

and

$$p_t + \langle \mathbf{q}, \nabla \rangle p + \rho a^2 \langle \nabla, \mathbf{q} \rangle = 0, \tag{5.1c}$$

where $a = a(\rho, p)$ is the local speed of sound. The assumption that the gas is polytropic leads to the entropic equation of state

$$p = K\rho^{\gamma} \tag{5.2}$$

in which the coefficient K depends on the entropy and γ is the ratio of specific heat, which for air is taken to be 1.4. For such a gas

$$a^2 = \frac{\gamma p}{\rho}. \tag{5.3}$$

These quasilinear equations form a hyperbolic system and are called Euler's equations. For some simplicity in the general theory, we took $\mathbf{u} = 0$ to be a basic solution of (2.1) and hence we write the equations of motion by replacing the density ρ by $\rho_* + \rho$, ($\rho_* =$ constant), pressure p by $p_* + p$ ($p_* =$ constant) and velocity $\mathbf{q}$ by $\mathbf{q}_* + \mathbf{q}$, where ($\mathbf{q}_* = 0$, $\rho = \rho_*$, $p = p_*$) represents the medium at rest and in uniform state and the symbols ρ, p and $\mathbf{q}$ now represent the perturbations. The equations of motion can be written in the form (2.1) where

$$\mathbf{u} = (\rho, q_1, q_2, q_3, p)^T, \quad A = I = \text{identity matrix.}$$

and

$$B_*^{(\alpha)} = \begin{bmatrix} q_\alpha & (\rho_* + \rho)\delta_{1\alpha} & (\rho_* + \rho)\delta_{2\alpha} & (\rho_* + \rho)\delta_{3\alpha} & 0 \\ 0 & q_\alpha & 0 & 0 & \frac{1}{\rho_*}\delta_{1\alpha} \\ 0 & 0 & q_\alpha & 0 & \frac{1}{\rho_*}\delta_{2\alpha} \\ 0 & 0 & 0 & q_\alpha & \frac{1}{\rho_*}\delta_{3\alpha} \\ 0 & \rho_* a_*^2 \left(1 + \frac{\rho}{\rho_*}\right)^{\gamma-1} \delta_{1\alpha} & \rho_* a_*^2 \left(1 + \frac{\rho}{\rho_*}\right)^{\gamma-1} \delta_{2\alpha} & \rho_* a_*^2 \left(1 + \frac{\rho}{\rho_*}\right)^{\gamma-1} \delta_{3\alpha} & q_\alpha \end{bmatrix},$$

$$\alpha = 1, 2, 3 \tag{5.4}$$

where δ_{ij} is the *Kronecker delta*, a_* is the value of local velocity of sound in the medium at rest. For the wave corresponding to the eigenvalue $\mathbf{q}+\mathbf{n}a$ the eikonal equation is

$$Q \equiv \phi_t + a_*\left(1+\frac{\rho}{\rho_*}\right)^{(\gamma-1)/2} |\nabla\phi| + \langle \mathbf{q}, \nabla\phi\rangle = 0, \tag{5.5}$$

and the corresponding right eigenvector is $\mathbf{r} = (r_1, r_2, r_3, r_4, r_5)^T$ where

$$r_1 = \frac{\rho_*}{a_*}\left(1+\frac{\rho}{\rho_*}\right)^{-(\gamma-3)/2}, \quad r_2 = n_1, \quad r_3 = n_2, \quad r_4 = n_3,$$
$$r_5 = \rho_* a_*\left(1+\frac{\rho}{\rho_*}\right)^{(\gamma+1)/2}. \tag{5.6}$$

In order that $\mathbf{l}A\mathbf{r} = 1$ we choose the left eigenvector as

$$l_1 = 0, \quad l_2 = \frac{1}{2}n_1, \quad l_3 = \frac{1}{2}n_2, \quad l_4 = \frac{1}{2}n_3, \quad l_5 = \frac{1}{2\rho_* a_*}\left(1+\frac{\rho}{\rho_*}\right)^{-(\gamma+1)/2}, \tag{5.7}$$

where $\mathbf{n} = (n_1, n_2, n_3)$ is the unit normal. The ray equations are given by

$$\frac{\mathrm{d}x_\alpha}{\mathrm{d}t} = q_\alpha + n_\alpha a_*\left(1+\frac{\rho}{\rho_*}\right)^{(\gamma-1)/2} = \chi_\alpha, \quad \text{say} \tag{5.8}$$

and

$$\frac{\mathrm{d}n_\alpha}{\mathrm{d}t} = -a_* L_\alpha\left(1+\frac{\rho}{\rho_*}\right)^{(\gamma-1)/2} - \sum_{\beta=1}^{3} n_\beta L_\alpha q_\beta = \Psi_\alpha, \tag{5.9}$$

where

$$\mathbf{L} \equiv (L_1, L_2, L_3) = \nabla - \mathbf{n}\langle \mathbf{n}, \nabla\rangle. \tag{5.10}$$

The expressions (2.18) and (3.15) to (3.19) when evaluated lead to the following set of equations of WNLRT up to order ϵ

$$\rho = \epsilon\frac{\rho_*}{a_*}w, \quad q_\alpha = \epsilon n_\alpha w, \quad p = \epsilon\rho_* a_* w$$

$$\frac{\mathrm{d}x_\alpha}{\mathrm{d}t} = n_\alpha\left(a_* + \epsilon\frac{\gamma+1}{2}w\right) \tag{5.11}$$

$$\frac{\mathrm{d}n_\alpha}{\mathrm{d}t} = -\epsilon\frac{\gamma+1}{2}L_\alpha w \tag{5.12}$$

and

$$\frac{\mathrm{d}w}{\mathrm{d}t} = \bar{\Omega}w, \tag{5.13}$$

where

$$\bar{\Omega} = -\frac{1}{2}a_*\langle\nabla, \mathbf{n}\rangle \tag{5.14}$$

is the mean curvature of the nonlinear wavefront and

$$\frac{\mathrm{d}}{\mathrm{d}t} = \frac{\partial}{\partial t} + \left(a_* + \epsilon\frac{\gamma+1}{2}w\right)\langle \mathbf{n}, \nabla\rangle \tag{5.15}$$

is the time rate of change along the rays given by (5.11) and (5.12). These are the same equations as derived in [15] where w is ϵw in the above equations. Since $|\mathbf{n}| = 1$, only two of the three equations (5.12) are independent. Therefore, the equations (5.11)–(5.13) form a system of 6 coupled equations for the determination of successive positions $\mathbf{x}$ of a nonlinear wavefront, the unit normal $\mathbf{n}$ and the wavefront intensity w. In the linear theory, $\mathbf{w}$ drops out of the (5.11) and (5.12) so that the ray equations decouple from the amplitude equation (5.13). In this case the rays and the successive positions of the wavefront can be constructed without any reference to the amplitude of the wave. This corresponds to the statement of Huygens' wavefront construction for the propagation of a linear wavefront. In our weakly nonlinear theory, the amplitude is related to the curvature of the wavefront (or the ray tube area) by the equation (5.13). The nonlinear rays stretch due to the presence of w in (5.11) and the wavefront rotates due to a non-uniform distribution of the amplitude on the wavefront (represented by $\mathbf{L}w$ in (5.12)). Thus the amplitude of the wave modifies the rays and the wavefront geometry which in turn effects the growth and decay of the amplitude.

Further we note that only the tangential derivatives, on a wavefront Ω_t at a time t, of w and n_α appear on the right hand side of the equations of WNLRT. Therefore, given the initial position Ω_0 of the wavefront and the distribution of the amplitude on it, all quantities on the right hand side of the equations (5.11)–(5.13) can be completely determined at $t = 0$ as in the case of a non-characteristic Cauchy problem. Hence, the evolution of the wavefront and the distribution of amplitude on it can be determined from these equations. This implies that, in the short wave approximation, the nonlinear wavefront is self propagating. The result is true not only for a compressible medium but for any continuum medium governed by the hyperbolic system. Huygens' method of wavefront construction has now been very elegantly extended to the construction of a nonlinear wavefront in the short wave limit.

As mentioned at the end of the last section, we shall now derive the shock ray theory for a weak shock from equations (5.11)–(5.15). Consider a weak shock wave propagating into a polytropic gas at rest ahead of it. Assume the shock also to belong to the characteristic field with eigenvalue $\mathbf{q} + \mathbf{n}a$, then shock will be followed by a one parameter family of nonlinear waves governed by the equations (5.11)–(5.13). Each one of these waves will catch up with the shock, interact with it and then disappear. A nonlinear wave while interacting with the shock will be instantaneously coincident with it in the short wave approximation considered in this paper. On the nonlinear wavefront, the transport equation (5.13) remains valid. Now we use the theorem ([14], p. 74).

Theorem. *For a weak shock, the shock ray velocity components are equal to the mean of the bicharacteristic velocity components just ahead and just behind the shock, provided we take the wavefront generating the characteristic surface to be instantaneously coincident with the shock surface. Similarly, the rate of turning of the shock front is equal to the mean of the rates of turning of such wavefronts just ahead and just behind the shock.*

We denote the unit normal to the shock front by $\mathbf{N}$. For the linear wavefront just ahead of the shock and instantaneously coincident with it (this is actually a linear wavefront moving with the ray velocity **N** multiplied by the local sound velocity a_*) $w = 0$ and the bicharacteristic velocity is $\mathbf{N}a_*$. For the nonlinear wavefront just behind the shock and instantaneously coincident with it, we denote the amplitude w by μ. Then μ is the shock amplitude of the weak shock under consideration. Using the theorem and the results (5.11) and (5.12) with $n = \mathbf{N}$, we get for a point **X** on the shock ray

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}T} = \frac{1}{2}\left\{a_*\mathbf{N} + \mathbf{N}\left(a_* + \epsilon\frac{\gamma+1}{2}\mu\right)\right\} = \mathbf{N}\left(a_* + \epsilon\frac{\gamma+1}{4}\mu\right) \tag{5.16}$$

$$\frac{\mathrm{d}\mathbf{N}}{\mathrm{d}T} = -\frac{1}{2}\left\{0 + \epsilon\frac{\gamma+1}{2}\mathbf{L}\mu\right\} = -\epsilon\frac{\gamma+1}{4}\mathbf{L}\mu, \tag{5.17}$$

where T is the time measured while moving along a shock ray. We take $w = \mu$ and $\mathbf{n} = \mathbf{N}$ in (5.13) and write it as

$$\begin{aligned}\frac{\mathrm{d}\mu}{\mathrm{d}T} &\equiv \left\{\frac{\partial}{\partial t} + \left(a_* + \epsilon\frac{\gamma+1}{4}\mu\right)\langle\mathbf{N},\nabla\rangle\right\}\mu \\ &= -\frac{1}{2}a_*\langle\nabla,\mathbf{N}\rangle\mu - \epsilon\,\frac{\gamma+1}{4}\mu\langle\mathbf{N},\nabla\rangle w,\end{aligned} \tag{5.18}$$

where we note that since μ is defined only on the shock front (and also on instantaneously coincident nonlinear wavefront behind it but not the other members of the one parameter family of wavefronts following it), the normal derivative $\langle\mathbf{N},\nabla\rangle\mu$ does not make sense mathematically. We introduce a new variable, defined on the shock

$$\mu_1 = \epsilon\langle\mathbf{n},\nabla\rangle w, \quad \text{on the shock front} \tag{5.19}$$

where ϵ appears to make $\mu_1 = O(1)$ since we wish to consider variation of w on a length scale over which the fast variable θ varies.

Equation (5.18) leads to the first compatibility condition along a shock ray

$$\frac{\mathrm{d}\mu}{\mathrm{d}T} = \bar{\Omega}_s\mu - \frac{\gamma+1}{4}\mu\mu_1, \tag{5.20}$$

where

$$\bar{\Omega}_s = -\frac{1}{2}a_*\langle\nabla,\mathbf{N}\rangle$$

is the value of $\bar{\Omega}$ for the nonlinear wavefront instantaneously coincident with the shock from behind.

To find the second compatibility condition along a shock, we differentiate (5.13) in the direction of **n** but on the length scale over which θ varies. On this length scale, $\mathbf{n}, \bar{\Omega}$ are constants and we get after rearranging some terms

$$\begin{aligned}&\left\{\frac{\partial}{\partial t} + \left(a_* + \epsilon\frac{\gamma+1}{4}w\right)\langle\mathbf{n},\nabla\rangle\right\}\langle\mathbf{n},\nabla\rangle w = -\frac{1}{2}\,a_*\langle\nabla,\mathbf{n}\rangle\langle\mathbf{n},\nabla\rangle w \\ &\quad - \epsilon\frac{\gamma+1}{4}w\langle\mathbf{n},\nabla\rangle^2 w - \epsilon\frac{\gamma+1}{4}\{\langle\mathbf{n},\nabla\rangle w\}^2.\end{aligned} \tag{5.21}$$

Writing this equation for the wavefront instantaneously coincident with the shock, multiplying it by ϵ and introducing a variable μ_2 by

$$\mu_2 = \epsilon^2\langle\mathbf{n},\nabla\rangle^2 w, \quad \text{on the shock} \tag{5.22}$$

we get

$$\frac{\mathrm{d}\mu_1}{\mathrm{d}T} = \bar{\Omega}_s\mu_1 - \frac{\gamma+1}{4}\mu_1^2 - \frac{\gamma+1}{4}\mu\mu_2 \tag{5.23}$$

which is the second compatibility condition along shock rays given by (5.16) and (5.17).

Similarly, higher order compatibility conditions can be derived. Thus, for the Euler's equations, we have derived the infinite system of compatibility conditions for a weak shock just from the dominant terms of our WNLRT (see [1], pp. 85–87).

As we have already mentioned, the shock ray theory is an exact theory (weak shock assumption is another independent assumption) but it is impossible to use it for computation for shock propagation. Prasad and Ravindran proposed a new theory of shock dynamics (NTSD) in 1990–91 (see [16]) according to which the system of equations (5.16), (5.17), (5.20) and (5.23) can be closed by dropping the term containing μ_2 from the equations (5.23). The NTSD has been found to be computationally very efficient and gives results which agree well with theoretical results (whatever available), experiment results and results obtained from computation of full gas-dynamics equations ([8] and a number of papers from Prasad, Ravindran and their collaborators). This new theory of shock dynamics forms the basis of extensive numerical computation by Monica and Prasad [9] to find the nonlinear effects in the linear caustic region.

Acknowledgement

The theory presented in § 2 was developed in collaboration with Prof. J K Hunter. The author sincerely thanks him for his valuable contribution.

References

[1] Anile M A, Hunter J K, Pantano P and Russo G, *Ray Methods for Nonlinear Waves in Fluids and Plasmas*, (Longman: 1993)

[2] Brio M and Hunter J K, Mach reflection for two-dimensional Burgers equation, *Physica* **D60** (1992) 194–207

[3] Hunter J K and Brio M, Weak shock reflection, *J. Fluid Mechanics* **410** (2000) 235–261

[4] Choquet-Bruhat Y, Ondes asymptotique et approachées pour systèmes nonlineares dèquations aux dérivées partielles nonlinéaires, *J. Math. Pure et Appl.* **48** (1969) 117–158

[5] Hunter J K, Asymptotic equations for nonlinear hyperbolic waves, Surveys in Applied Mathematics, (eds) J B Keller, W McLaughlin and C Papanicolaou (1995) (Plenum Press) vol. II

[6] Hunter J K, Nonlinear wave diffraction, Geometrical Optics and Related Topics (eds.) F Colombini, and N Lerner (1997) (Berkhauser) pp. 221–243

[7] Keller J B, Geometrical acoustics, I: Theory of weak shock waves, *J. Appl. Phys.* **25** (1954) 938–947

[8] Kevlahan NK-R, The propagation of weak shocks in non-uniform flow, *J. Fluid Mech.* **327** (1996) 167–197

[9] Monica A and Phoolan Prasad, Propagation of a curved weak shock, communicated for publication in *J. Fluid Mech.*

[10] Parker D F, Nonlinearity, relaxation and diffusion in acoustic and ultrasonics, *J. Fluid Mech.* **39** (1969) 793–815

[11] Parker D F, An asymptotic theory for oscillatory nonlinear signals, *J. Inst. Math. Appli.* **7** (1971) 92–110

[12] Phoolan Prasad, Approximation of the perturbation equations of a quasilinear hyperbolic system in the neighbourhood of a bicharacteristic, *J. Math. Anal. Appl.* **50** (1975) 470–482

[13] Phoolan Prasad, Extension of Huyghen's construction of a wavefront to a non-linear wavefront and a shock-front, *Curr. Sci.* **56** (1987) 50–54

[14] Phoolan Prasad, Propagation of a Curved Shock and Nonlinear Ray Theory, *Pitman Researches in Mathematics Series, No. 292* (1993) (Longman)

[15] Phoolan Prasad, A nonlinear ray theory, *Wave Motion* **20** (1994) 21–31

[16] Prasad P and Ravindran R, A new theory of shock dynamics, Part II: Numerical results, *Appl. Math. Lett.* **3** 107–109

[17] Phoolan Prasad and Renuka Ravindran, A theory of nonlinear waves in multi-dimensions with special reference to surface-water-waves, *J. Inst. Math. Appl.* **20** (1977) 9–20

[18] Phoolan Prasad and Sangeeta K, Numercial simulation of converging nonlinear wavefronts, *J. Fluid Mech.* **385** (1999) 1–20

[19] Renuka Ravindran and Phoolan Prasad, Kinematics of a shock front and resolution of a hyperbolic caustic (a review article), *Advances in nonlinear waves* (London: Pitman) (1985) vol. II, pp. 77–99

[20] Sturtevant B and Kulkarny V A, The focusing of weak shock waves, *J. Fluid Mech.* **73** (1976) 651–671

[21] Tabak E and Rosales R R, Weak shock focusing and von-Neumann paradox of oblique shock reflection, *Phys. Fluids* **6** (1994) 1874–1892

[22] Whitham G B, On the propagation of weak shock waves, *J. Fluid Mech.* **1** (1956) 290–318

Proc. Indian Acad. Sci. (Math. Sci.), Vol. 110, No. 4, November 2000, pp. 449–465.
© Printed in India

Steady-state response of a micropolar generalized thermoelastic half-space to the moving mechanical/thermal loads

RAJNEESH KUMAR and SUNITA DESWAL

Department of Mathematics, Kurukshetra University, Kurukshetra 136 119, India

MS received 14 January 2000

Abstract. Microrotation effect of a load applied normal to the boundary and moving at a constant velocity along one of the co-ordinate axis in a generalized thermoelastic half-space is studied. The analytical expressions of the displacement component, force stress, couple stress and temperature field for two different theories i.e. Lord-Shulman (L-S) and Green-Lindsay (G-L) for supersonic, subsonic and transonic velocities in case of mechanical and thermal sources applied, are obtained by the use of Fourier transform technique. The integral transforms have been inverted by using a numerical technique and the numerical results are illustrated graphically for magnesium crystal-like material.

Keywords. Steady-state response; Galilean transformation; Mach numbers; micropolar; generalized thermoelasticity.

1. Introduction

The classical theory of heat conduction predicts infinite speed of heat transportation, if a material conducting heat is subjected to a thermal disturbance, which contradicts the physical facts. During the last three decades non-classical theories have been developed to remove this paradox. Lord and Shulman [12] incorporated a flux rate term into the Fourier's law of heat conduction and formulated a generalized theory admitting finite speed for thermal signals. Green and Lindsay [8] have developed a temperature rate dependent thermoelasticity by including temperature rate among the constitutive variables which does not violate the classical Fourier's law of heat conduction when the body under consideration has a centre of symmetry and this theory also predicts a finite speed of heat propagation. These theories consider heat propagation as a wave phenomenon rather than a diffusion phenomenon. In view of the experimental evidence available in favour of finiteness of heat propagation speed, generalized thermoelasticity theories are supposed to be more realistic than the conventional theory in dealing with practical problems involving very large heat fluxes and or short time intervals, like those occurring in laser units and energy channels.

Modern engineering structures are often made up of materials possessing an internal structure. Polycrystalline materials and materials with fibrous or coarse grain structure come in this category. Classical elasticity is inadequate to represent the behaviour of such materials. The analysis of such materials requires incorporating the theory of oriented media. 'Micropolar elasticity' termed by Eringen [3] is used to describe deformation of elastic media with oriented particles. A micropolar continuum is a collection of interconnected particles in the form of small rigid bodies undergoing both translational and rotational motions. Typical examples of such materials are granular media and multimolecular bodies, whose microstructures act as an evident part in their macroscopic responses. The physical nature of these materials needs an asymmetric description of deformation, while theories for classical continua fail to accurately predict their physical and mechanical behaviour. For this reason, micropolar theories were developed by Eringen [3–5] for elastic solids, fluids and further for non-local polar fields and are now universally accepted.

The linear theory of micropolar thermoelasticity was developed by extending the theory of micropolar continua to include thermal effects by Eringen [2] and Nowacki [13]. Steady state response to moving loads in elasticity have been discussed in Fung [7]. Different authors [1, 9, 10, 11, 14, 16–18] discussed different problems in *elasticity/ micropolar elasticity/micropolar thermoelasticity*. Following [2, 8, 12], we study the disturbance due to a moving load on the surface $z = 0$ of the micropolar generalized thermoelastic half-space by applying integral transform technique.

2. Formulation and solution of the problem

We consider a micropolar generalized thermoelastic solid occupying the half space in an undisturbed state and initially at uniform temperature T_0. The rectangular Cartesian co-ordinates are introduced having origin on the surface $z = 0$ and z-axis pointing vertically into the medium. Let us consider a pressure pulse $P(x + Ut)$, which is moving with a constant speed in the negative x-direction for an infinitely long time so that a steady state prevails in the neighbourhood of the loading as seen by an observer moving with the load.

Following Eringen [2], Lord and Shulman [12] and Green and Lindsay [8], the field equations and stress-strain temperature relations in micropolar generalized thermoelastic solid without body forces, body couples and heat sources can be written as

$$(\lambda+\mu)\nabla(\nabla\cdot\vec{u})+(\mu+K)\nabla^2\vec{u}+K\nabla x\vec{\phi}-\nu\left(1+t_1\frac{\partial}{\partial t}\right)\nabla T=\rho\frac{\partial^2\vec{u}}{\partial t^2}, \qquad (1)$$

$$(\alpha+\beta+\gamma)\nabla(\nabla\cdot\vec{\phi})-\gamma\nabla x(\nabla x\vec{\phi})+K\nabla x\vec{u}-2K\vec{\phi}=\rho j\frac{\partial^2\vec{\phi}}{\partial t^2}, \qquad (2)$$

$$K^*\nabla^2T=\rho C^*\left(\frac{\partial T}{\partial t}+t_0\frac{\partial^2T}{\partial t^2}\right)+\nu T_0\left(\frac{\partial}{\partial t}+\delta_{1k}t_0\frac{\partial^2}{\partial t^2}\right)\nabla\cdot\vec{u} \qquad (3)$$

and

$$t_{ij}=\lambda u_{r,r}\delta_{ij}+\mu(u_{i,j}+u_{j,i})+K(u_{j,i}-\in_{ijr}\phi_r)-\nu\left(T+t_1\frac{\partial T}{\partial t}\right)\delta_{ij}, \qquad (4)$$

$$m_{ij}=\alpha\phi_{r,r}\delta_{ij}+\beta\phi_{i,j}+\gamma\phi_{j,i}, \qquad (5)$$

where $\lambda, \mu, K, \alpha, \beta, \gamma$ are material constants, ρ is the density, j is the microinertia, K^* is the coefficient of thermal conductivity, $\nu = (3\lambda + 2\mu + K)\alpha_t$, α_t is the coefficient of linear thermal expansion, C^* is the specific heat at constant strain, $T(x, z, t)$ is the change in temperature of the medium at any time; t_0, t_1 are the thermal relaxation times, $\vec{u}$ is the displacement vector and $\vec{\phi}$ is the microrotation vector. For Lord-Shulman (L-S) theory $t_1 = 0, k = 1$ and for Green-Lindsay (G-L) theory $t_1 > 0$ and $k = 2$. The thermal relaxations t_0 and t_1 satisfy the inequality $t_1 \geq t_0 \geq 0$ for the G-L theory only.

For two dimensional problem, we assume

$$\vec{u}=(u_x,0,u_z), \qquad \vec{\phi}=(0,\phi_2,0) \qquad (6)$$

and introducing potential functions q and ψ defined by

$$u_x = \frac{\partial q}{\partial x} + \frac{\partial \psi}{\partial z}, \qquad u_z = \frac{\partial q}{\partial z} - \frac{\partial \psi}{\partial x}, \qquad \text{where } \psi = (-\vec{U})_y, \tag{7}$$

in equations (1)–(3), we obtain

$$\left[\nabla^2 - \frac{\rho}{(\lambda+2\mu+K)}\frac{\partial^2}{\partial t^2}\right] q - \frac{\nu}{(\lambda+2\mu+K)}\left(1 + t_1\frac{\partial}{\partial t}\right) T = 0, \tag{8}$$

$$\left[\nabla^2 - \frac{\rho}{(\mu+K)}\frac{\partial^2}{\partial t^2}\right]\psi - \frac{K}{(\mu+K)}\phi_2 = 0, \tag{9}$$

$$\left[\nabla^2 - \frac{2K}{\gamma} - \frac{\rho j}{\gamma}\frac{\partial^2}{\partial t^2}\right]\phi_2 + \frac{K}{\gamma}\nabla^2\psi = 0, \tag{10}$$

$$\left[\nabla^2 - \frac{\rho C^*}{K^*}\left(\frac{\partial}{\partial t} + t_0\frac{\partial^2}{\partial t^2}\right)\right] T - \frac{\nu T_0}{K^*}\left(\frac{\partial}{\partial t} + \delta_{1k}t_0\frac{\partial^2}{\partial t^2}\right)\nabla^2 q = 0. \tag{11}$$

Following Fung [7], a Galilean transformation

$$x^* = x + Ut, \quad z^* = z, \qquad t^* = t \tag{12}$$

is introduced, then the boundary conditions would be independent of t^* and assuming the dimensionless variables defined by the expressions

$$x' = \frac{\omega^*}{c_2}x^*, \quad z' = \frac{\omega^*}{c_2}z^*, \quad t' = \omega^* t^*, \quad t_1' = \omega^* t_1, \quad t_0' = \omega^* t_0, \quad T' = \frac{T}{T_0},$$
$$q' = \frac{\rho\omega^{*2}}{\nu T_0}q, \quad \psi' = \frac{\rho\omega^{*2}}{\nu T_0}\psi, \quad \phi_2' = \frac{\rho c_2^2}{\nu T_0}\phi_2, \quad t_{ij}' = \frac{t_{ij}}{\nu T_0}, \quad m_{ij}' = \frac{\omega^*}{c_2\nu T_0}m_{ij}, \tag{13}$$

where

$$\omega^* = \frac{\rho C^* c_2^2}{K^*}, \qquad c_2^2 = \frac{\mu}{\rho},$$

in equations (8)–(11), we get (after suppressing the primes)

$$\left[\left(1 - \frac{U^2}{c_1^2}\right)\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2}\right] q + a_2\left(1 + t_1\frac{U}{c_2}\frac{\partial}{\partial x}\right) T = 0, \tag{14}$$

$$\left[\left(1 - \frac{U^2}{c_3^2}\right)\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2}\right]\psi - a_3\phi_2 = 0, \tag{15}$$

$$\left[\left(1 - \frac{U^2}{c_4^2}\right)\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2} - 2a_1\right]\phi_2 + a_1\left(\frac{\partial^2}{\partial x^2} + \frac{\partial}{\partial z^2}\right)\psi = 0 \tag{16}$$

$$\left[\left(1 - \frac{U^2}{c_5^2}\right)\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2} - \frac{U}{c_2}\frac{\partial}{\partial x}\right] T - \in \left[\frac{U}{c_2}\frac{\partial}{\partial x} + \delta_{1k}t_0\frac{U^2}{c_2^2}\frac{\partial^2}{\partial x^2}\right]\left[\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2}\right] q = 0, \tag{17}$$

where

$$c_1^2 = \frac{\lambda+2\mu+K}{\rho}, \quad c_3^2 = \frac{\mu+K}{\rho}, \quad c_4^2 = \frac{\gamma}{\rho j}, \quad c_5^2 = \frac{K^*}{t_0\rho C^*},$$
$$a_1 = \frac{Kc_2^2}{\gamma\omega^{*2}}, \quad a_2 = \frac{\rho c_2^2}{\lambda+2\mu+K}, \quad a_3 = \frac{K}{\mu+K}, \quad \in = \frac{\nu^2 T_0}{\rho\omega^* K^*} \tag{18}$$

We define the Fourier transform as

$$\hat{f}(\xi, z) = \int_{-\infty}^{\infty} f(x, z) e^{i\xi x} dx \tag{19a}$$

and its inverse by

$$f(x, z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(\xi, z) e^{-i\xi x} d\xi \tag{19b}$$

Applying the Fourier transform defined by (19a) on eqs (14)–(17), we obtain

$$\left[\frac{d^2}{dz^2} - \xi^2\left(1 - \frac{U^2}{c_1^2}\right)\right]\hat{q} - a_2 d_2 \hat{T} = 0, \tag{20}$$

$$\left[\frac{d^2}{dz^2} - \xi^2\left(1 - \frac{U^2}{c_3^2}\right)\right]\hat{\psi} - a_3 \hat{\phi}_2 = 0, \tag{21}$$

$$\left[\frac{d^2}{dz^2} - \xi^2\left(1 - \frac{U^2}{c_4^2}\right) - 2a_1\right]\hat{\phi}_2 + a_1\left(\frac{d^2}{dz^2} - \xi^2\right)\hat{\psi} = 0, \tag{22}$$

$$\left[\frac{d^2}{dz^2} - \xi^2\left(1 - \frac{U^2}{c_5^2}\right) + i\xi U_1\right]\hat{T} + \in d_1\left(\frac{d^2}{dz^2} - \xi^2\right)\hat{q} = 0, \tag{23}$$

where

$$d_1 = U_1 i\xi + \delta_{1k} t_0 U_1^2 \xi^2, \quad d_2 = 1 - i\xi t_1 U_1, \quad U_1 = U/c_2. \tag{24}$$

We introduce the match numbers

$$M_i = U/c_i$$

and the parameters α_i and α_i' as

$$\alpha_i = \sqrt{1 - M_i^2}, \quad \text{if } M_i < 1; \quad \alpha_i' = \sqrt{M_i^2 - 1}, \quad \text{if } M_i > 1, \quad i = 1, 3, 4, 5 \tag{25}$$

in eqs (20)–(23) and then eliminate $\hat{T}$ and $\hat{\phi}_2$ from the resulting expressions to obtain the following equations

$$\left[\frac{d^4}{dz^4} + A \frac{d^2}{dz^2} + B\right][\hat{q}] = 0 \tag{26}$$

and

$$\left[\frac{d^4}{dz^4} + D \frac{d^2}{dz^2} + E\right][\hat{\psi}] = 0, \tag{27}$$

where

$$A = -\xi^2(\alpha_1^2 + \alpha_5^2) + i\xi U_1 + \epsilon a_2 d_1 d_2, \quad B = \xi^2[\alpha_1^2 \alpha_5^2 \xi^2 - i\xi U_1 \alpha_1^2 - \epsilon a_2 d_1 d_2],$$
$$D = -\xi^2(\alpha_3^2 + \alpha_4^2) + a_1(a_3 - 2), \quad E = \xi^2[\alpha_3^2 \alpha_4^2 \xi^2 + 2a_1 \alpha_3^2 - a_1 a_3]. \tag{28}$$

The solutions of eqs (26) and (27), satisfying the radiation conditions are

$$\hat{q} = A_1 \exp(-\xi_1 z) + A_2 \exp(-\xi_2 z), \tag{29}$$

$$\hat{T} = q_1 A_1 \exp(-\xi_1 z) + q_2 A_2 \exp(-\xi_2 z), \tag{30}$$

$$\hat{\psi} = A_3 \exp(-\xi_3 z) + A_4 \exp(-\xi_4 z), \tag{31}$$

$$\hat{\phi}_2 = q_3 A_3 \exp(-\xi_3 z) + q_4 A_4 \exp(-\xi_4 z), \tag{32}$$

where $\xi_{1,2}^2$ and $\xi_{3,4}^2$ are roots of eqs (26) and (27) respectively and are given by

$$\xi_{1,2}^2 = [-A \pm \sqrt{A^2 - 4B}]/2, \qquad \xi_{3,4}^2 = [-D \pm \sqrt{D^2 - 4E}]/2, \tag{33}$$

and

$$q_{1,2} = \frac{1}{a_2 d_2}[\xi_{1,2}^2 - \alpha_1^2 \xi^2], \qquad q_{3,4} = \frac{1}{a_3}[\xi_{3,4}^2 - \alpha_3^2 \xi^2]. \tag{34}$$

Case 1. *Mechanical source acting on the surface*

In moving co-ordinates, the boundary conditions are

$$t_{zz} = -P\delta(x), \quad t_{zx} = m_{zy} = \frac{\partial T}{\partial z} = 0 \qquad \text{at} \quad z = 0, \tag{35}$$

where $\delta(x)$ is Dirac-delta function in moving co-ordinates.

Sub-case 1.1: *Subsonic.* $M_i < 1 (i = 1, 3, 4, 5)$. Making use of eqs (4), (5), (7), (12) and (13) in the boundary conditions (35); applying the transform defined by (19a) and substituting the values of $\hat{q}, \hat{T}, \hat{\psi}$ and $\hat{\phi}_2$ from equations (29)–(32) in the resulting expressions, we obtain the expressions for displacement components, stresses and temperature field as

$$\hat{u}_x = -\frac{1}{\Delta}[i\xi(\Delta_1 e^{-\xi_1 z} + \Delta_2 e^{-\xi_2 z}) + \xi_3 \Delta_3 e^{-\xi_3 z} + \xi_4 \Delta_4 e^{-\xi_4 z}], \tag{36}$$

$$\hat{u}_z = -\frac{1}{\Delta}[\xi_1 \Delta_1 e^{-\xi_1 z} + \xi_2 \Delta_2 e^{-\xi_2 z} - i\xi(\Delta_3 e^{-\xi_3 z} + \Delta_4 e^{-\xi_4 z})], \tag{37}$$

$$\hat{t}_{zz} = \frac{1}{\Delta}[f_1 \Delta_1 e^{-\xi_1 z} + f_2 \Delta_2 e^{-\xi_2 z} - i b_2 \xi(\xi_3 \Delta_3 e^{-\xi_3 z} + \xi_4 \Delta_4 e^{-\xi_4 z})], \tag{38}$$

$$\hat{t}_{zx} = \frac{1}{\Delta}[i\xi b_2(\xi_1 \Delta_1 e^{-\xi_1 z} + \xi_2 \Delta_2 e^{-\xi_2 z}) + f_3 \Delta_3 e^{-\xi_3 z} + f_4 \Delta_4 e^{-\xi_4 z}], \tag{39}$$

$$\hat{m}_{zy} = \frac{-b_7}{\Delta}[\xi_3 q_3 \Delta_3 e^{-\xi_3 z} + \xi_4 q_4 \Delta_4 e^{-\xi_4 z}], \tag{40}$$

$$\hat{T} = \frac{1}{\Delta}[q_1 \Delta_1 e^{-\xi_1 z} + q_2 \Delta_2 e^{-\xi_2 z}], \tag{41}$$

where

$$\begin{aligned}\Delta = {} & q_1 \xi_1 [\xi_3 q_3 (f_2 f_4 - b_2^2 \xi^2 \xi_2 \xi_4) - \xi_4 q_4 (f_2 f_3 - b_2^2 \xi^2 \xi_2 \xi_3)] \\ & + q_2 \xi_2 [-\xi_3 q_3 (f_1 f_4 - b_2^2 \xi^2 \xi_1 \xi_4) + \xi_4 q_4 (f_1 f_3 - b_2^2 \xi^2 \xi_1 \xi_3)],\end{aligned}$$

$$\Delta_1 = P\xi_2 q_2 (f_4 \xi_3 q_3 - f_3 \xi_4 q_4), \quad \Delta_2 = P\xi_1 q_1 (f_3 \xi_4 q_4 - f_4 \xi_3 q_3),$$

$$\Delta_3 = P\xi_4 q_4 b_2 \xi \xi_2 \xi_1 (q_1 - q_2), \quad \Delta_4 = P\xi_3 q_3 b_2 \xi \xi_1 \xi_2 (q_2 - q_1),$$

$$f_i = b_3 \xi_i^2 - b_1 \xi^2 - d_2 q_i, \quad (i = 1, 2); \quad f_j = b_6 \xi_j^2 + b_4 \xi^2 - b_5 q_j, \quad (j = 3, 4),$$

$$b_1 = \frac{\lambda}{\rho c_2^2}, \quad b_2 = \frac{(2\mu + K)}{\rho c_2^2}, \quad b_3 = b_1 + b_2, \quad b_4 = \frac{\mu}{\rho c_2^2},$$
$$b_5 = \frac{K}{\rho c_2^2}, \quad b_6 = b_4 + b_5, \quad b_7 = \frac{\gamma \omega^{*2}}{\rho c_2^4}. \tag{42}$$

Particular case. Neglecting microrotational effect i.e $(\alpha = \beta = \gamma = K = j = 0)$ in eqs (36)–(41), the expressions for displacement components, force stresses and temperature field are obtained in a thermoelastic medium as

$$\hat{u}_x = -\frac{1}{\Delta_0}[i\xi(\Delta_1' \mathrm{e}^{-\xi_1 z} + \Delta_2' \mathrm{e}^{-\xi_2 z}) + \xi_3' \Delta_3' \mathrm{e}^{-\xi_3' z}], \tag{43}$$

$$\hat{u}_z = -\frac{1}{\Delta_0}[\xi_1 \Delta_1' \mathrm{e}^{-\xi_1 z} + \xi_2 \Delta_2' \mathrm{e}^{-\xi_2 z} - i\xi \Delta_3' \mathrm{e}^{-\xi_3' z}], \tag{44}$$

$$\hat{t}_{zz} = \frac{1}{\Delta_0}[f_1' \Delta_1' \mathrm{e}^{-\xi_1 z} + f_2' \Delta_2' \mathrm{e}^{-\xi_2 z} - ib_2' \xi \xi_3' \Delta_3' \mathrm{e}^{-\xi_3' z}], \tag{45}$$

$$\hat{t}_{zx} = \frac{1}{\Delta_0}[i\xi b_2'(\xi_1 \Delta_1' \mathrm{e}^{-\xi_1 z} + \xi_2 \Delta_2' \mathrm{e}^{-\xi_2 z}) + f_3' \Delta_3' \mathrm{e}^{-\xi_3' z}], \tag{46}$$

$$\hat{T} = \frac{1}{\Delta_0}[q_1 \Delta_1' \mathrm{e}^{-\xi_1 z} + q_2 \Delta_2' \mathrm{e}^{-\xi_2 z}], \tag{47}$$

where

$$\Delta_0 = -f_3'[f_1' \xi_2 q_2 - f_2' \xi_1 q_1], \quad \Delta_1' = Pf_3' \xi_2 q_2,$$
$$\Delta_2' = -Pf_3' \xi_1 q_1, \quad \Delta_3' = P\xi b_2 \xi_1 \xi_2 (q_2 - q_1),$$
$$f_i' = b_3' \xi_i^2 - b_1 \xi^2 - d_2 q_i, \quad (i = 1, 2); \quad f_3' = b_4(\xi_3'^2 + \xi^2),$$
$$b_2' = 2\mu / \rho c_2^2, \quad b_3' = b_1 + b_2', \quad \xi_3'^2 = \alpha_3^{0^2} \xi^2, \quad \alpha_3^0 = \sqrt{1 - U_1^2}$$

and α_1 in the expressions of A, B and $q_{1,2}$ takes the form

$$\alpha_1^0 = \sqrt{1 - M_1^{0^2}}, \quad M_1^0 = \frac{U}{c_1^0}, \quad c_1^{0^2} = (\lambda + 2\mu)/\rho. \tag{48}$$

Sub-case 1.2: *Supersonic.* $M_i > 1$ $(i = 1, 3, 4, 5)$. In this case, $A, B, D, E, q_{1,2}$ and $q_{3,4}$ in the expressions (36)–(41), take the form

$$A = \xi^2(\alpha_1'^2 + \alpha_5'^2) + i\xi U_1 + a_2 d_1 d_2 \epsilon,$$
$$B = \xi^2(\alpha_1'^2 \alpha_5'^2 \xi^2 + i\xi U_1 \alpha_1'^2 - a_2 d_1 d_2 \epsilon),$$
$$D = +\xi^2(\alpha_3'^2 + \alpha_4'^2) + a_1(a_3 - 2), \quad E = \xi^2[\alpha_3'^2 \alpha_4'^2 \xi^2 - 2a_1 \alpha_3'^2 - a_1 a_3],$$
$$q_{1,2} = \frac{1}{a_2 d_2}(\xi_{1,2}^2 + \alpha_1'^2 \xi^2), \quad q_{3,4} = \frac{1}{a_3}(\xi_{3,4}^2 + \alpha_3'^2 \xi^2). \tag{49}$$

Particular case. If we neglect microrotational effect, then the expressions (43)–(47) are obtained in a thermoelastic medium with A, B and $q_{1,2}$ defined by equation (49), with $\alpha_1^{0^2}$ and $\alpha_3^{0^2}$ replaced by $\alpha_1^{*^2}$ and $\alpha_3^{*^2}$ respectively, where

$$\alpha_1^* = \sqrt{M_1^{0^2} - 1}, \qquad \alpha_3^* = \sqrt{U_1^2 - 1}. \tag{50}$$

Sub-case 1.3: *Transonic.* $M_{1,3} < 1$, $M_{4,5} > 1$. In this case A, B, D, E in the expressions (36)–(41), take the form

$$\begin{aligned}
&A = \xi^2(-\alpha_1^2 + \alpha_5'^2) + i\xi U_1 + a_2 d_2 d_1 \epsilon,\\
&B = -\xi^2[\alpha_1^2\alpha_5'^2\xi^2 + i\xi U_1\alpha_1^2 + a_2 d_2 d_1 \epsilon],\\
&D = \xi^2(\alpha_4'^2 - \alpha_3^2) + a_1(a_3 - 2), \quad E = \xi^2(-\alpha_3^2\alpha_4'^2\xi^2 + 2a_1\alpha_3^2 - a_1 a_3). \quad (51)
\end{aligned}$$

Particular case. Neglecting microrotational effect, the analytical expressions for displacement components, force stresses and temperature field are again given by eqs (43)–(47) with A and B defined by (51) and with α_1^2 replaced by $\alpha_1^{0^2}$ defined by (48).

Case 2. *Thermal source acting on the surface*

The boundary conditions in this case are

$$t_{zz} = t_{zx} = m_{zy} = 0, \quad \frac{\partial T}{\partial z} = P\delta(x) \quad \text{at } z = 0. \qquad (52)$$

Sub-case 2.1: *Subsonic.* $M_i < 1$ $(i = 1, 3, 4, 5)$. Following the procedure, adopted in Case 1, and using the boundary conditions (52), the expressions for displacement components, stresses and temperature field are again given by eqs (36)–(41) with $\Delta_i\ (i = i, \ldots, 4)$ replaced by Δ_i^* given by the following equations.

$$\begin{aligned}
&\Delta_1^* = P[-\xi_3 q_3(f_2 f_4 - \xi^2 b_2^2 \xi_2 \xi_4) + \xi_4 q_4(f_2 f_3 - \xi^2 b_2^2 \xi_2 \xi_3)],\\
&\Delta_2^* = P[\xi_3 q_3(f_1 f_4 - \xi^2 b_2^2 \xi_1 \xi_4) - \xi_4 q_4(f_1 f_3 - \xi^2 b_2^2 \xi_1 \xi_3)],\\
&\Delta_3^* = P\xi_4 q_4 b_2 \xi(f_2\xi_1 - f_1\xi_2), \quad \Delta_4^* = P\xi_3 q_3 b_2 \xi(f_1\xi_2 - f_2\xi_1). \qquad (53)
\end{aligned}$$

Particular case. In a thermoelastic medium; displacement components, force stresses and temperature field are given by expressions (43)–(47) with $\Delta_i'\ (i = 1, 2, 3)$ replaced by Δ_i^0, where

$$\begin{aligned}
&\Delta_1^0 = -P(f_2' f_3' - \xi^2 b_2'^2 \xi_2 \xi_3'), \quad \Delta_2^0 = P(f_1' f_3' - \xi^2 b_2'^2 \xi_1 \xi_3'),\\
&\Delta_3^0 = P\xi b_2'(f_1'\xi_2 - f_2'\xi_1). \qquad (54)
\end{aligned}$$

Sub-case 2.2: *Supersonic.* $M_i > 1$ $(i = 1, 3, 4, 5)$. In the eqs (36)–(41), if we define A, B, D, E and $q_i\ (i = 1, \ldots, 4)$ by (49) and replace $\Delta_i\ (i = 1, \ldots, 4)$ by Δ_i^* given by (53), we obtain the expressions for displacement components, force stresses and temperature field for this case.

Particular case. In the absence of microrotational effect, the expressions (43)–(47) are obtained in a thermoelastic medium with $\Delta_1'\ (i = 1, 2, 3)$ replaced by Δ_i^0 given by (54), with $A, B, q_{1,2}$ defined by (49) and with $\alpha_{1,3}^{0^2}$ replaced with $\alpha_{1,3}^{*^2}$ defined by (50).

Sub-case 2.3: *Transonic.* $M_{1,3} < 1$, $M_{4,5} > 1$. In this case A, B, D, E in the expressions (36)–(41) are given by (51) and $\Delta_i\ (i = 1, \ldots, 4)$ are replaced by Δ_i^* given by (53).

Particular case. If microrotational effect is neglected, the expressions for displacement components, force stresses and temperature field are given by equations (43)–(47), with $\Delta_1'(i = 1, 2, 3)$ replaced by Δ_i^0 given by eq. (54), with A and B defined by (51) and with α_1^2 replaced by α_1^{02} defined by (48).

Special cases.

I. For L-S theory, d_1 and d_2 in the expressions (36)–(41) for all the cases become as

$$d_1 = U_1 i\xi + t_0 U_1^2 \xi^2, \quad d_2 = 1. \tag{55}$$

II. In G-L theory, for each case, d_1 in the expressions (36)–(41), takes the form

$$d_1 = U_1 i\xi. \tag{56}$$

3. Inversion of the transforms

To obtain the solution of the problem in the physical domain, we must invert the transforms in (36)–(41) for the two theories in case of subsonic, supersonic and transonic velocities for both the cases. These expressions are functions of z and the parameter of Fourier transform ξ and hence are of the form $\hat{f}(\xi, z)$. To get the function $f(x, z)$ in the physical domain, we invert the Fourier transform using

$$f(x,z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-i\xi x}\hat{f}(\xi,z)\mathrm{d}\xi = \frac{1}{\pi}\int_{0}^{\infty}(\cos(\xi x)f_{\mathrm{e}} - i\sin(\xi x)f_{\mathrm{o}})\mathrm{d}\xi, \tag{57}$$

where f_{e} and f_{o} are respectively even and odd parts of the function $\hat{f}(\xi, z)$. The method for evaluating this integral is described by Press *et al* [15], which involves the use of Romberg's integration with adaptive step size. This, also uses the results from successive refinements of the extended trapezoidal rule followed by extrapolation of the results to the limit when the step size tends to zero.

4. Numerical results and discussion

We take the magnesium crystal as an example for the purpose of numerical evaluation, the physical data for which is given as [6]

$$\rho = 1.74\,\mathrm{gm/cm^3}, \quad j = 0.2 \times 10^{-15}\,\mathrm{cm^2}, \quad \lambda = 9.4 \times 10^{11}\,\mathrm{dyne/cm^2},$$
$$\mu = 4.0 \times 10^{11}\,\mathrm{dyne/cm^2}, \quad K = 1.0 \times 10^{11}\,\mathrm{dyne/cm^2}, \quad \gamma = 0.779 \times 10^{-4}\,\mathrm{dyne},$$
$$K^* = 0.6 \times 10^{-2}\,\mathrm{cal/cm\,sec\,^\circ C}, \quad C^* = 0.23\,\mathrm{cal/gm\,^\circ C},$$
$$\epsilon = 0.073, \quad T_0 = 23^\circ\mathrm{C}, \quad t_0 = 6.131 \times 10^{-13}\,\mathrm{sec}, \quad t_1 = 8.765 \times 10^{-13}\,\mathrm{sec}.$$

The variations of normal displacement $U_z (= u_z/P)$, normal force stress $T_{zz} (= t_{zz}/P)$, tangential couple stress $M_{zy} (= m_{zy}/P)$ and temperature field $T^* (= T/P)$ with distance x for L-S and G-L theories have been shown by (a) solid line (—) and solid line with centered symbols (—×—×—×—) respectively for micropolar generalized thermoelastic (MGTE) medium and (b) dashed line (– –) and dashed line with centered symbols (⊖ ⊖ ⊖) respectively for generalized thermoelastic (GTE) medium. These variations for subsonic, supersonic and transonic velocities in case of mechanical and thermal sources applied are shown in figures 1–24.

4.1 *Discussion for Case* 1

4.1.1 *Mechanical source*; *subsonic.* Due to microrotation effect the values of normal displacement in MGTE medium are large in the range $0 \leq x \leq 2.0$ and $4.5 \leq x \leq 8.0$; small in the range of $2.0 < x < 4.5$ and $8.0 < x \leq 10.0$ in comparison to GTE medium

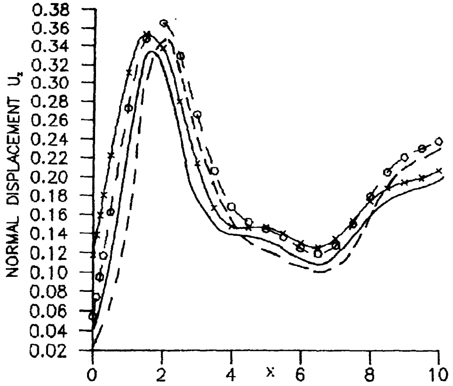


Figure 1. Variations of normal displacement $U_Z(= u_Z/P)$ with distance x (mechanical source; subsonic).

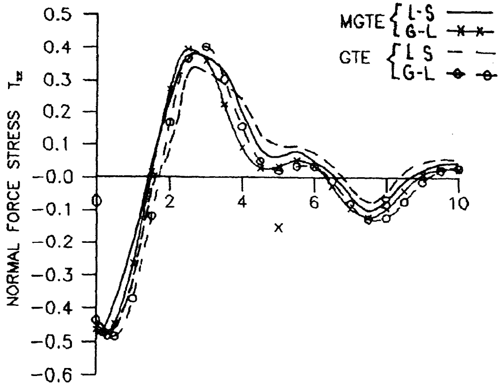


Figure 2. Variations of normal force stress $T_{zz}(= t_{zz}/P)$ with distance x (mechanical source; subsonic).

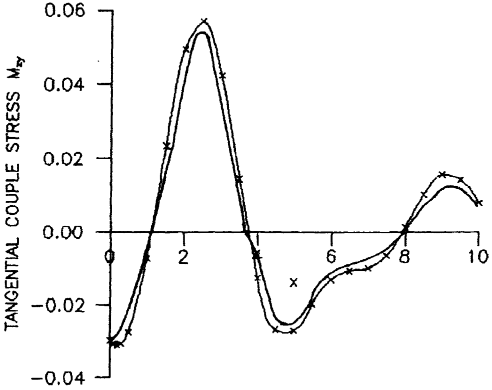


Figure 3. Variations of tangential couple stress $M_{zy}(= m_{zy}/P)$ with distance x (mechanical source; subsonic).

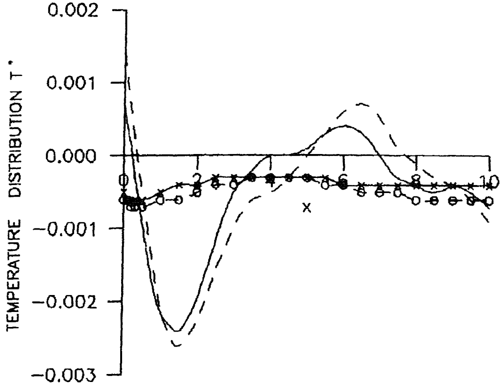


Figure 4. Variations of temperature distribution $T^*(= T/P)$ with distance x (mechanical source; subsonic).

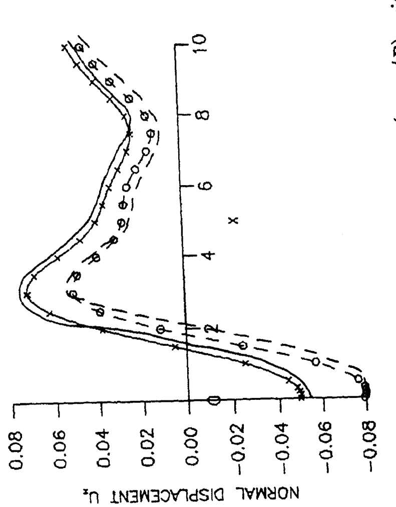


Figure 5. Variations of normal displacement $U_z(= u_z/P)$ with distance x (mechanical source; supersonic).

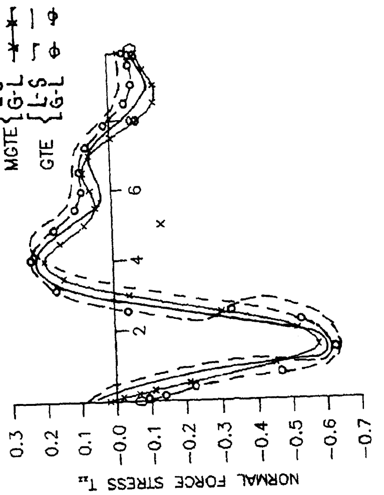


Figure 6. Variations of normal force stress $T_{zz}(= t_{zz}/P)$ with distance x (mechanical source; supersonic).

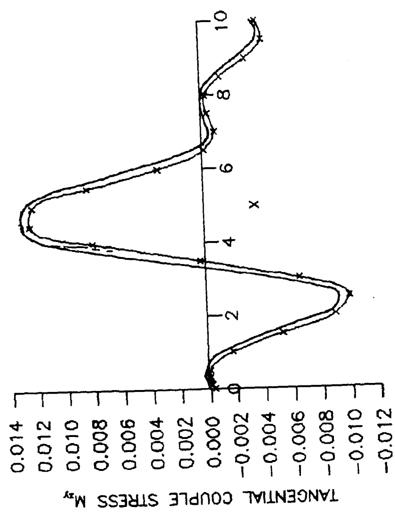


Figure 7. Variations of tangential couple stress $M_{zy}(= m_{zy}/P)$ with distance x (mechanical source; supersonic).

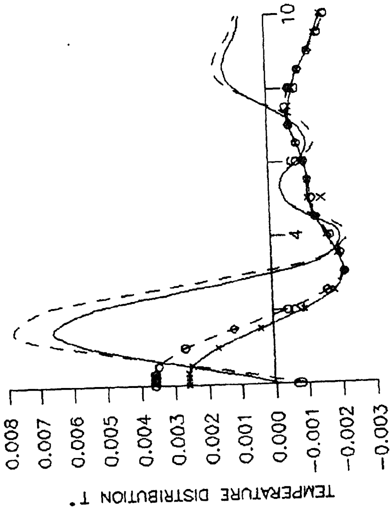


Figure 8. Variations of temperature distribution $T^*(= T/P)$ with distance x (mechanical source; supersonic).

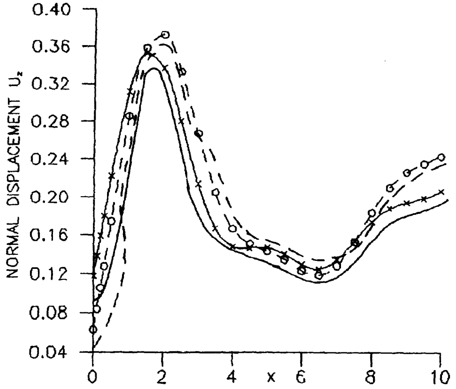


Figure 9. Variations of normal displacement $U_z(=u_z/P)$ with distance x (mechanical source; transonic).

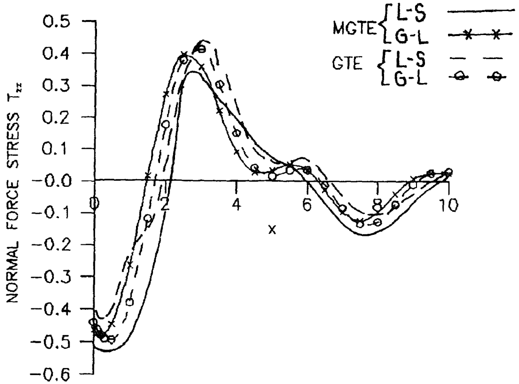


Figure 10. Variations of normal force stress $T_{zz}(=t_{zz}/P)$ with distance x (mechanical source; transonic).

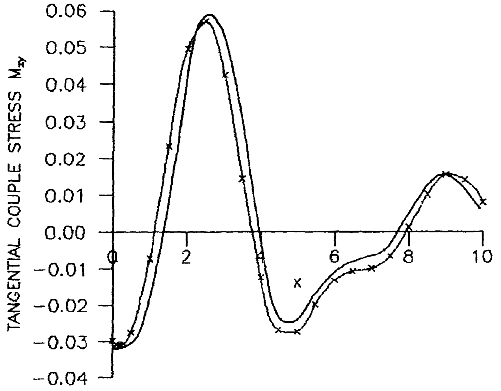


Figure 11. Variations of tangential couple stress $M_{zy}(=m_{zy}/P)$ with distance x (mechanical source; transonic).

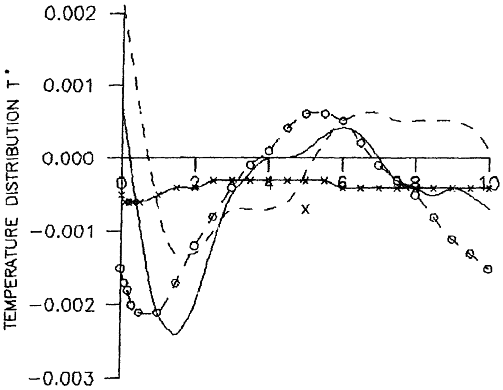


Figure 12. Variations of temperature distribution $T^*(=T/P)$ with distance x (mechanical source; transonic).

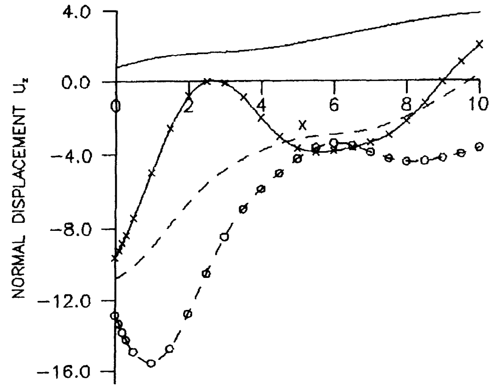


Figure 13. Variations of normal displacement $U_z(= u_z/P)$ with distance x (thermal source; subsonic).

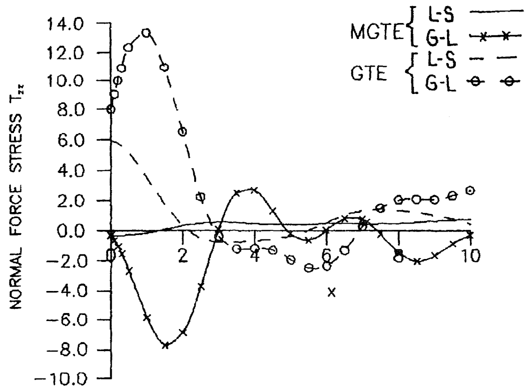


Figure 14. Variations of normal force stress $T_{zz}(= t_{zz}/P)$ with distance x (thermal source; subsonic).

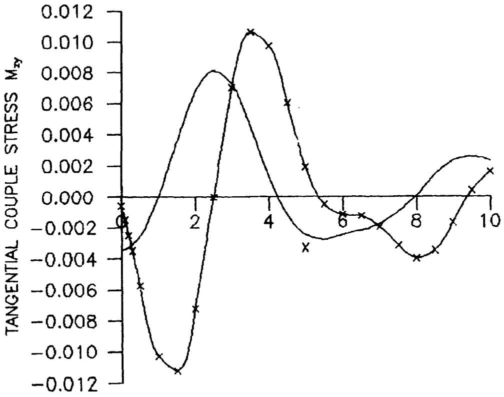


Figure 15. Variations of tangential couple stress $M_{zy}(= m_{zy}/P)$ with distance x (thermal source; subsonic).

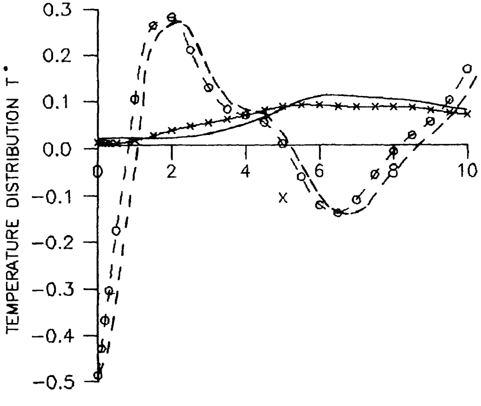


Figure 16. Variations of temperature distribution $T^*(= T/P)$ with distance x (thermal source; subsonic).

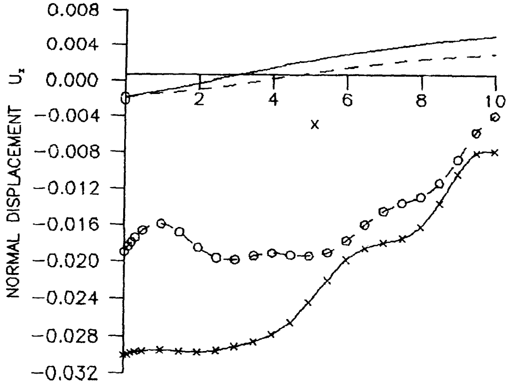


Figure 17. Variations of normal displacement $U_z(= u_z/P)$ with distance x (thermal source; supersonic).

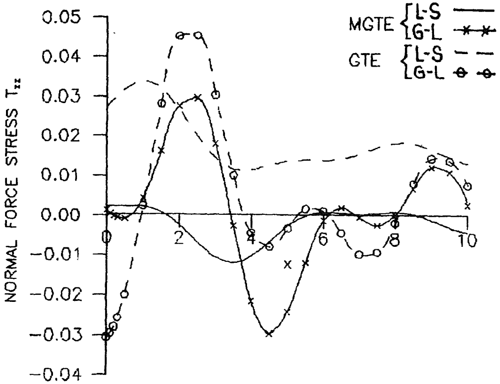


Figure 18. Variations of normal force stress $T_{zz}(= t_{zz}/P)$ with distance x (thermal source; supersonic).

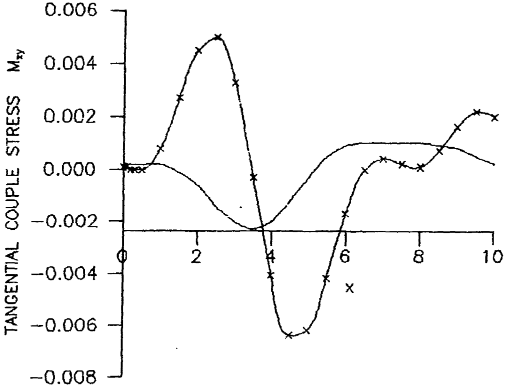


Figure 19. Variations of tangential couple stress $M_{zy}(= m_{zy}/P)$ with distance x (thermal source; supersonic).

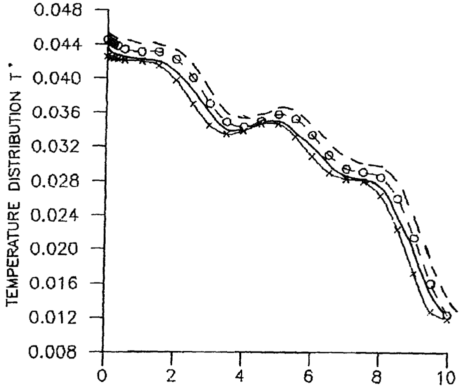


Figure 20. Variations of temperature distribution $T^*(= T/P)$ with distance x (thermal source; supersonic).

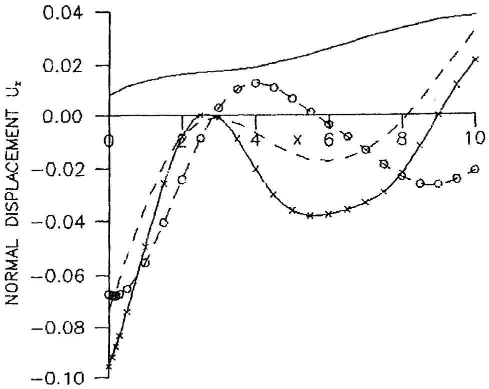


Figure 21. Variations of normal displacement $U_z(= u_z/P)$ with distance x (thermal source; transonic).

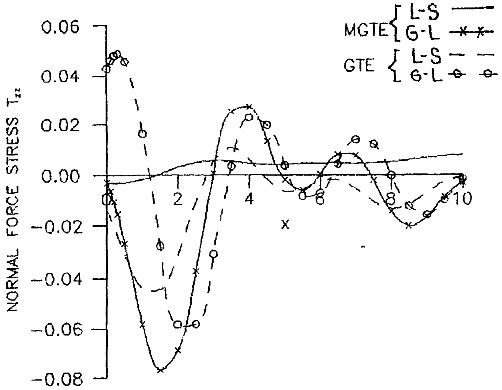


Figure 22. Variations of normal force stress $T_{zz}(= t_{zz}/P)$ with distance x (thermal source; transonic).

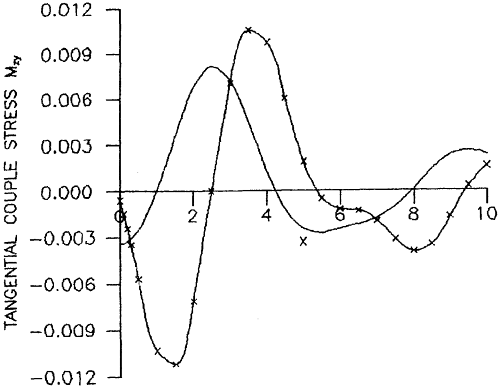


Figure 23. Variations of tangential couple stress $M_{zy}(= m_{zy}/P)$ with distance x (thermal source; transonic).

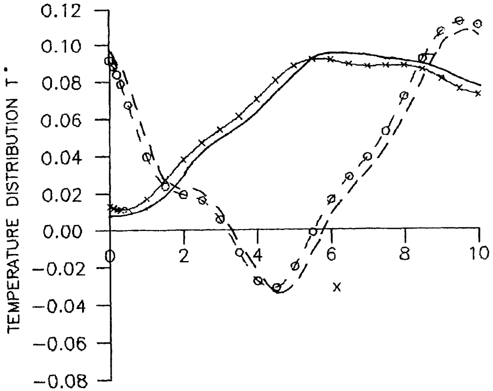


Figure 24. Variations of temperature distribution $T^*(= T/P)$ with distance x (thermal source; transonic).

for both the theories. Also, the values of normal displacement for L-S theory are small in comparison to G-L theory in both the media as shown in figure 1. Microrotation effect on normal force stress can be observed from figure 2, where, in case of L-S theory the values of normal force stress in MGTE medium are large in the range $0 \leq x \leq 3.5$ and small in the range $3.5 < x \leq 10.0$ in comparison to GTE medium. In case of G-L theory, values of normal force stress in MGTE medium are large in the range $0 \leq x \leq 2.5$, $5.0 \leq x \leq 6.0$ and $7.5 \leq x \leq 10.0$; small in the range $2.5 < x < 5.0$ and $6.0 < x < 7.5$ in comparison to GTE medium. The behaviour of the variations of tangential couple stress in MGTE medium for L-S and G-L theories is similar as shown in figure 3. The range of variations of temperature field in case of G-L theory is very small in comparison to L-S theory for both the media. Due to microrotation effect the values of temperature field in MGTE medium are small in the range $0 \leq x \leq 1.5$ and $5.5 \leq x \leq 9.0$; large in the range $1.5 < x < 5.5$ and $9.0 < x \leq 10.0$ in comparison to GTE medium in case of L-S theory, whereas for G-L theory values in MGTE medium are large in comparison to GTE medium in the range $0 \leq x \leq 10.0$, as shown in figure 4.

4.1.2 *Mechanical source*; *supersonic*. Due to microrotational effect, the values of normal displacement in MGTE medium are large in comparison to GTE medium for L-S and G-L theories as can be observed from figure 5. The values of normal force stress for L-S theory in MGTE medium are small in the range $0 \leq x \leq 1.0$ and $4.0 \leq x \leq 10.0$ but large in the range $1.0 < x < 4.0$ in comparison to GTE medium, whereas for G-L theory the values in MGTE medium are large in the range $0 \leq x \leq 2.5$ and small in the range $2.5 \leq x \leq 10.0$ in comparison to GTE medium as depicted in figure 6. Figure 7 illustrates the variations of tangential couple stress in MGTE medium, where, the values for L-S theory are large in comparison to G-L theory. In case of L-S theory the values of temperature field in MGTE medium are large in the range $0 \leq x \leq 1.0$ and $3.5 \leq x \leq 8.0$; small in the range $1.0 < x < 3.5$ and $8.0 < x \leq 10.0$ in comparison to GTE medium, whereas in case of G-L theory the values in MGTE medium are small in the range $0 \leq x \leq 3.0$ and large in the range $3.0 < x \leq 10.0$ in comparison to GTE medium as shown in figure 8. It is also observed that the range of variation of temperature field for G-L theory is very small in comparison to L-S theory where, in figure 8 the variations for G-L theory have been shown after multiplying the original values by 10 in both the media.

4.1.3 *Mechanical source*; *transonic*. In case of L-S theory the values of normal displacement for MGTE medium are large in the initial range $0 \leq x \leq 1.0$ and remain small in the further range $1.0 < x \leq 10.0$ in comparison to GTE medium, whereas for G-L theory the values in MGTE medium are large in the range $0 \leq x \leq 1.5$ and $5.0 \leq x \leq 7.5$; small in the range $1.5 < x < 5.0$ and $7.5 < x \leq 10.0$ in comparison to GTE medium, as can be observed from figure 9. Due to microrotation effect the values of normal force stress in MGTE medium are small in comparison to GTE medium in case of L-S theory, whereas in case of G-L theory the values in MGTE medium are large in the range $0 \leq x \leq 2.5$ and $5.0 \leq x \leq 10.0$ but small in the range $2.5 < x < 5.0$ in comparison to GTE medium as illustrated in figure 10. The values of tangential couple stress in MGTE medium for L-S theory are small in the range $0 \leq x \leq 2.5$ and $9.0 \leq x \leq 10.0$ but large in the range $2.5 < x < 9.0$ in comparison to G-L theory as seen in figure 11. The range of variations of temperature field for L-S theory lie in a large range in comparison to G-L theory for both the media as observed from figure 12, where, the original values of temperature field in GTE medium for L-S theory have been divided by 10. Due to microrotation effect the values of temperature field in MGTE medium vary in a small range in comparison to GTE medium for L-S and G-L theories.

4.2 *Discussion for Case* 2

4.2.1 *Thermal source*; *subsonic.* Due to microrotation effect, the values of normal displacement and normal force stress in MGTE medium lie in a very small range in comparison to GTE medium for both L-S and G-L theories, where, the variations for normal displacement and normal force stress for L-S and G-L theories in MGTE medium have been shown after multiplying the original values by 10^2. These variations of normal displacement and normal force stress have been shown in figures 13 and 14 respectively. The values of tangential couple stress in MGTE medium for L-S theory are large in the range $0 \leq x \leq 0.5$ and $3.0 \leq x \leq 7.0$; small in the range $0.5 < x < 3.0$ and $7.0 < x \leq 10.0$ in comparison to G-L theory. Also the variations for L-S theory lie in a small range in comparison to G-L theory as can be observed from figure 15, where, the original values of tangential couple stress for L-S theory have been multiplied by 10 to depict the comparison. Due to microrotation, the variations of temperature field in MGTE medium lie in a very small range in comparison to GTE medium for both the theories, as is evident from figure 16.

4.2.2 *Thermal source*; *supersonic.* The values of normal displacement in MGTE medium are large for L-S theory and small for G-L theory in comparison to GTE medium as can be noticed from figure 17. Also, the values of normal displacement for L-S theory are large in comparison to G-L theory in both the media. The range of variations of normal force stress for L-S theory is very small in comparison to G-L theory for both the media as is clear from figure 18, where, the variations for L-S theory in both the media are shown after multiplying the original values by 10. For L-S theory, due to microrotation effect the values of normal force stress in MGTE medium are small in comparison to GTE medium, whereas for G-L theory the values in MGTE medium are large in the range $0 \leq x \leq 1.0$ and $6.0 \leq x \leq 8.0$; small in the range $1.0 < x < 6.0$ and $8.0 < x \leq 10.0$ in comparison to GTE medium. The variations of tangential couple stress for L-S theory lie in a very small range in comparison to G-L theory, so the original values of tangential couple stress for L-S theory have been multiplied by 10 to depict the comparison. Microrotation effect on temperature field for L-S and G-L theories is shown in figure 20, where, the values of temperature field in MGTE medium are small in comparison to GTE medium for both the theories. Also, the values of temperature field for L-S theory are large in comparison to G-L theory in both the media.

4.2.3 *Thermal source*; *transonic.* Due to microrotation effect the values of normal displacement for L-S theory in MGTE medium are large in comparison to GTE medium, whereas for G-L theory variations of normal displacement in MGTE medium lie in a small range in comparison to GTE medium as illustrated in figure 21, where, in GTE medium the variations for G-L theory are shown after dividing the original values by 10. The range of variation of normal force stress in MGTE medium is small in comparison to GTE medium for L-S and G-L theories as is evident from figure 22, where the original values for G-L theory in GTE medium have been divided by 10 to depict the variations simultaneously. Since the values of tangential couple stress in MGTE medium for L-S theory lie in a small range in comparison to G-L theory, therefore, the original values for L-S theory have been multiplied by 10 to illustrate the comparison in figure 23. Due to microrotation effect the values of temperature field in MGTE medium are small in the range $0 \leq x \leq 1.5$ and $8.0 \leq x \leq 10.0$; large in the range $1.5 < x < 8.0$ in comparison to GTE medium for L-S and G-L theories, as is evident from figure 24.

Thus, from the above numerical results, we conclude that microrotation has a significant effect on normal displacement, normal force stress and temperature field for both the theories in all the cases. In case of mechanical source applied, microrotation effect is more appreciable for temperature field in comparison to normal displacement and normal force stress, whereas in case of thermal source applied, microrotation effect is more significant for normal displacement and normal force stress in comparison to temperature field.

References

[1] Dhaliwal R S, The steady-state axisymmetric problem of micropolar thermoelasticity, *Arch. Mech.* **23** (1971) 705–714

[2] Eringen A C, Foundation of micropolar thermoelasticity, Course of lectures no. 23, (CISM Udine, Springer) (1970)

[3] Eringen A C, Linear theory of micropolar elasticity, *J. Math. Mech.* **15** (1966) 909–923

[4] Eringen A C, Theory of micropolar fluids, *J. Math. Mech.* **16** (1966) 1–18

[5] Eringen A C, Non-local polar field theories, in: Continuum Physics (ed.) A C Eringen (New York: Academic Press) (1976) vol. IV, 205–267

[6] Eringen A C, Plane waves in non-local micropolar elasticity, *Int. J. Engg. Sci.* **22** (1984) 1113–1121

[7] Fung Y C, *Foundation of Solid Mechanics* (New Delhi: Prentice- Hall) (1968)

[8] Green A E and Lindsay K A, Thermoelasticity, *J. Elasticity* **2** (1972) 1–5

[9] Kumar R, Chadha T K and Debnath L, Lamb's plane problem in micropolar thermoelastic medium with stretch, *Int. J. Math. and Math. Sci.* **10** (1987) 187–198

[10] Kumar R and Singh B, Wave propagation in a micropolar generalized thermoelastic body with stretch, *Proc. Indian Acad. Sci.* **106** (1996) 183–189

[11] Kumar R and Gogna M L, Steady-state response to moving loads in micropolar elastic medium with stretch, *Int. J. Engg. Sci.* **30** (1992) 811–820

[12] Lord H W and Shulman Y, A generalized dynamical theory of thermoelasticity, *J. Mech. Phys. Solids* **15** (1967) 299–306

[13] Nowacki W, Couple stresses in the theory of thermoelasticity. *Proc. IUTAM Symposia* (Vienna: Springer-Verlag) (1966) 259–278

[14] Nath S and Sengupta P R, Steady-state response to moving loads in an elastic solid media, *Indian J. Pure Appl. Math.* **30** (1999) 317–327

[15] Press W H, Teukolsky S A, Vellerling W T and Flannery B P, *Numerical Recipes* (Cambridge: Cambridge University Press) (1986)

[16] Shanker M U and Dhaliwal R S, Dynamic coupled thermoelastic problems in micropolar theory – I, *Int. J. Engg. Sci.* **13** (1975) 121–128

[17] Sengupta P R and Ghosh B, Waves and vibrations in micropolar elastic medium I. Steady-state response to moving loads, *Arch. Mech.* **29** (1977) 273–287

[18] Singh B and Kumar R, Reflection of plane waves from the flat boundary of a micropolar generalized thermoelastic half space, *Int. J. Engg. Sci.* **36** (1998) 865–890

SUBJECT INDEX

AUTHOR INDEX

End of one hundred and tenth volume